

Undergraduate Lecture Notes in Physics

Alessandro Bettini

A Course in Classical Physics 3 — Electromagnetism



Undergraduate Lecture Notes in Physics

Undergraduate Lecture Notes in Physics (ULNP) publishes authoritative texts covering topics throughout pure and applied physics. Each title in the series is suitable as a basis for undergraduate instruction, typically containing practice problems, worked examples, chapter summaries, and suggestions for further reading.

ULNP titles must provide at least one of the following:

- An exceptionally clear and concise treatment of a standard undergraduate subject.
- A solid undergraduate-level introduction to a graduate, advanced, or non-standard subject.
- A novel perspective or an unusual approach to teaching a subject.

ULNP especially encourages new, original, and idiosyncratic approaches to physics teaching at the undergraduate level.

The purpose of ULNP is to provide intriguing, absorbing books that will continue to be the reader's preferred reference throughout their academic career.

Series editors

Neil Ashby
University of Colorado, Boulder, CO, USA

William Brantley
Department of Physics, Furman University, Greenville, SC, USA

Matthew Deady
Physics Program, Bard College, Annandale-on-Hudson, NY, USA

Michael Fowler
Department of Physics, University of Virginia, Charlottesville, VA, USA

Morten Hjorth-Jensen
Department of Physics, University of Oslo, Oslo, Norway

Michael Inglis
SUNY Suffolk County Community College, Long Island, NY, USA

Heinz Klose
Humboldt University, Oldenburg, Niedersachsen, Germany

Helmy Sherif
Department of Physics, University of Alberta, Edmonton, AB, Canada

More information about this series at <http://www.springer.com/series/8917>

Alessandro Bettini

A Course in Classical Physics

3 — Electromagnetism



Springer

Alessandro Bettini
Dipartimento di Fisica e Astronomia
Università di Padova
Padua
Italy

ISSN 2192-4791 ISSN 2192-4805 (electronic)
Undergraduate Lecture Notes in Physics
ISBN 978-3-319-40870-5 ISBN 978-3-319-40871-2 (eBook)
DOI 10.1007/978-3-319-40871-2

Library of Congress Control Number: 2016945945

© Springer International Publishing Switzerland 2016

The text is partially based on the book “Elettromagnetismo”, A. Bettini, Zanichelli, 1991. The author own all rights in the former publications.

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

This is the third in a series of four volumes, all written at an elementary calculus level. The complete course covers the most important areas of classical physics, such as mechanics, thermodynamics, statistical mechanics, electromagnetism, waves and optics. The volumes are the result of a translation, an in-depth revision, and an update of the Italian version published by Decibel–Zanichelli. This third volume deals with classical electromagnetism.

The vast majority of physics phenomena naturally taking place around us originate in electromagnetic forces. The largest portion of the technology that holds sway over so much of our lives and our civilization is the result of our knowledge and control of electromagnetic forces.

The forces that keep atoms together internally, binding electrons to nuclei, and externally, binding those atoms within molecules, are electromagnetic. As a consequence, the energies developed in all chemical reactions are electromagnetic, including the biochemical ones, which are the basis of life itself. Contact forces, such as friction, between solid surfaces are electromagnetic, and so is the drag force acting on a body moving in a fluid. Elastic forces, cohesion forces, and the force that results from the earth's magnetism are all electromagnetic. Light itself is an electromagnetic wave, whose wavelength is in the range to which our eyes are sensitive. The radio waves we use in telecommunications, radio, television, and cellular phones are electromagnetic as well, utilizing much greater wavelengths. As a matter of fact, all the phenomena on scales larger than those of the atomic and molecular have a gravitational or electromagnetic origin.

Modern technology is more than 99 % reliant on electromagnetism. In hydroelectric power stations, for example, big turbines are moved by water falling on them through large pipes under pressure. The turbines move electromagnetic generators, made of massive copper coils rotating between the poles of an electromagnet. The generators produce an electromotive force that is then distributed through a network of thousands of copper wires across distances of hundreds of kilometers to factories, offices, and houses. Here, the electric power is used by electric motors to produce all types of objects, to control chemical processes, or

simply to light our rooms or wash our dishes. All of these are electromagnetic processes. Cellular phones emit and receive electromagnetic waves, which are produced or detected and amplified by electronic circuits. Our computers store, process, and transmit information using electronic circuits of ever-increasing complexity and miniaturization.

However, the electromagnetic nature of natural phenomena does not appear at first sight and remained substantially unknown until roughly two centuries ago. Lighting is the phenomenon with the most evident (to us) electric nature, but it was not the genesis of the study of electric and magnetic phenomena. On the contrary, the first observations were in regard to the curious properties of amber, which, when rubbed, attracted small pieces of papyrus, and of magnetite, a stone capable of attracting pieces of iron. These phenomena were reported by the Greek philosopher Thales from Miletus in the sixth century BC. Twenty-two centuries had to pass before the first systematic observations and the first attempts at interpretation of electric and magnetic phenomena would occur, with William Gilbert's publication of his work in the book *De magnete* in 1600. Still, almost two more centuries would go by before Charles Augustin de Coulomb would make the fundamental measurement of the electric force between two charges and its dependence on their distance in 1785, finally paving the way for electric and magnetic research.

The basic reason for this late scientific birth of electromagnetism can be traced to the fact that, even if the electric force is billions of billions of billions of billions stronger than the gravitational force, it can be attractive or repulsive, depending on the sign of the charges, and because matter is made of positive and negative charges so exactly equal and opposite and so intimately mixed together that they perfectly balance one another. Nobody knows, even today, the reason for this perfect equality. Phenomena in different sectors of physics, mechanics or acoustics, for example, have always been known to the common man, and later scientists have gradually discovered the underlying laws. Contrastingly, the entire electromagnetic world is a discovery of science.

The life spans of the principal contributors to electromagnetism are shown in Fig. 1.

The next fundamental step forward after Coulomb was credited to Alessandro Volta, who published his discovery of the pile in 1800. The production of voltages and electric currents became available for further experiments and the pace of progress grew very rapidly, leading to a complete understanding of electromagnetism in less than a century. Volta's pile made possible the experiment with which Hans Christian Ørsted, in 1820, first discovered the magnetic effects of electric currents, connecting electricity and magnetism for the first time. Between 1820 and 1826, André Marie Ampère completely clarified the relation between the magnetic field and electric currents with a series of beautiful experiments. In 1831, Michael Faraday discovered electromagnetic induction, the phenomenon in which magnetic fields variable with time produce electric fields. Finally, in 1865, James Clerk Maxwell wrote the differential equations that contain the complete theory of electromagnetism. The equations not only foresaw a new phenomenon, electromagnetic waves, but also that light itself is such a wave. The theoretical prediction

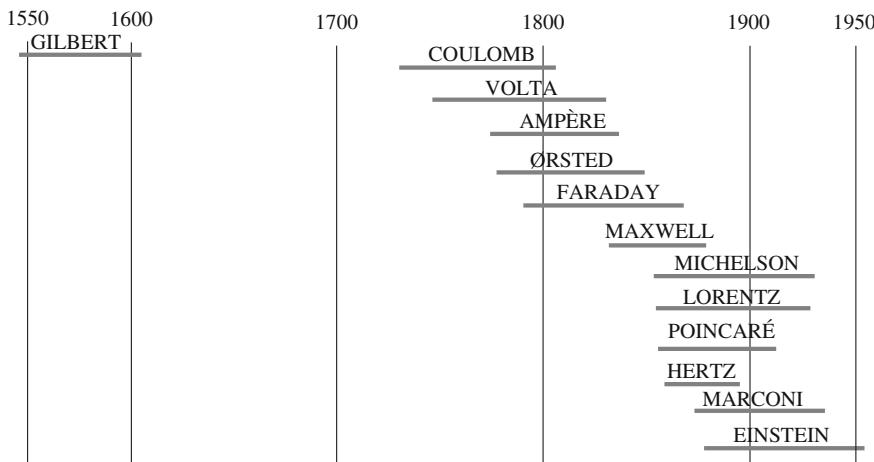


Fig. 1 Life spans of the greater contributors to electromagnetism

of the electromagnetic waves, specifically those that we call radio waves, was experimentally confirmed by Henrich Rudolf Hertz with a series of experiments, also quite beautiful, between 1886 and 1889. Twelve years later, in 1901, Guglielmo Marconi succeeded in sending the first radio transmission across the Atlantic Ocean. The message, consisting simply of the Morse code signal for the letter “s”, traveled more than 2000 miles from Cornwall in England to Newfoundland in Canada.

As opposed to Galilei–Newton mechanics, the electromagnetic theory also appeared to hold in its original formulation of Maxwell at the highest velocities up to the speed of light. In other words, the electromagnetic theory was born being already relativistically correct. Better still, it was that very progress in the in-depth experimental study of electromagnetic phenomena, in particular, with the outstanding experiment of Albert Abraham Michelson (with Morley) in 1887, and the revolutionary theoretical analysis between 1895 and 1905 by, mainly, Hendrik Antoon Lorentz, Henri Poincaré and Albert Einstein, that led to special relativity.

Classical electromagnetism is a magnificent scientific construction that quantitatively describes a huge number of phenomena and is the basis of modern technology. However, it does not work in the interpretation of electromagnetic interaction at atomic or smaller scales. The experimental and theoretical progress over the past century led to the development of quantum electrodynamics, which contains the classical electrodynamics as an approximation valid on large enough scales and fully explains the phenomena down to the smallest scales explored experimentally thus far.

The scope of these lectures is the description, at an introductory level, of classical electromagnetism. The reader is assumed to be acquainted with differential calculus, including the simplest partial differential equations, the gradient, divergence and curl operators and their basic theorems.

Physics is an experimental science, meaning that it is based on the experimental method, which was developed by Galileo Galilei in the seventeenth century. The process of understanding physical phenomena is not immediate, but rather, it advances by trial and error, in a series of experiments, which might lead, with a bit of fortune and a lot of thinking, to the discovery of the governing laws. Induction of the process of physical laws goes back from the observed effects to their causes, and, as such, cannot be purely logical. Once a physical law is found, it is necessary to consider all its possible consequences. This is now a deductive process, which is logical and similar to that of mathematics. Each of the consequences, the predictions, of the law must then be experimentally verified. If only one prediction is found to be false by the experiment, even if thousands of them had been found true, it is enough to prove that the law is false. This implies that we can never be completely sure that a law is true; indeed, the number of its possible predictions is limitless, and at any historical moment, not all of them have been controlled. However, this is the price we must pay in choosing the experimental method, which has allowed humankind to advance much further in the last four centuries than in all the preceding millennia.

The path of science is complex, laborious, and highly nonlinear. In its development, errors have been made and hypotheses have been advanced that turned out to be false, but ultimately, laws were discovered. The knowledge of at least a few of the most important aspects of this process is indispensable for developing the mental capabilities necessary for anybody who wishes to contribute to the progress of natural sciences, whether they pursue applications or teach them. It is for this reason that we shall read and discuss the descriptions some of these authors have put forth of their fundamental experiments.

Each chapter of the book starts with a brief introduction on a scope that will give the reader a preliminary idea of the arguments he/she will find. There is no need to fully understand these introductions at the first reading, as all the arguments are fully developed in the subsequent pages.

At the end of each chapter, the reader will find a summary and a number of queries with which to check his/her level of understanding of the chapter's arguments. The difficulty of the queries is variable; some of them are very simple, some more complex, a few are true numerical exercises. However, the book does not contain any sequence of full exercises, considering the existence of very good textbooks dedicated specifically to that.

The first four chapters deal with electrostatics, namely electric phenomena under time-independent conditions. Chapter 1 is on electrostatics in a vacuum. The concept of the electric charge is introduced and the basic properties of this fundamental physical quantity are discussed. We then discuss the force between charges at rest, introduce the concept of the electric field and discuss its properties. The materials can be schematically classified, from the electric point of view, into two main classes, the conductors and the insulators, which are also called dielectrics. We deal with the former in Chap. 2, and latter in Chap. 4. Chapter 3 is dedicated to the energy of the electrostatic systems in a vacuum and with conductors. Chapter 5 treats the steady electric currents. Chapter 6 is on magnetostatics,

namely magnetic phenomena under time-independent conditions, in a vacuum. In Chap. 7, we start our study in dynamic, namely time-dependent, situations with the important phenomena of electromagnetic induction, which link electricity and magnetism. Chapter 8 is dedicated to the study of the energy of the magnetostatic systems in a vacuum. In Chap. 9, we study magnetic phenomena in the presence of matter, in particular, diamagnetism, paramagnetism, and ferromagnetism. Finally, in Chap. 10 , we reach the full description of the Maxwell equations, both in a vacuum and in matter, discover the new phenomena they foresee and study the Lorentz invariance of the equations.

Padua, Italy

Alessandro Bettini

Acknowledgments

Thanks to Pavia University for the picture of the Alessandro Volta pile (reproduction by Gelside Guatterini 1999). Museo per la Storia dell'Università—Sistema Museale di Ateneo—Pavia University in Fig. 1.8.

Figure 2.7 is reproduced with permission of the Physics and Astronomy Department of Padua University.

Figure 9.25 is reprinted with permission from C.H. Fowler and E.M. Fryer; Physical Review **94** (1953) 52. Copyright (1953) by the American Physical Society.

Contents

1	Electrostatic Field in a Vacuum	1
1.1	Electric Charge	2
1.2	Coulomb's Law	7
1.3	The Electrostatic Field	13
1.4	Calculating Electric Fields	16
1.5	Electrostatic Potential	19
1.6	Generating Electrostatic Potential Differences	22
1.7	Calculating Electrostatic Potentials	24
1.8	Measuring the Elementary Charge. Millikan's Experiment	27
1.9	Invariance of the Electric Charge	32
1.10	The Solid Angle	33
1.11	The Flux of E and the Gauss Theorem	36
1.12	Graphic Representation of the Electric Field	39
1.13	Applications of the Gauss Law	41
1.14	Discontinuities of the Electric Field	48
1.15	Poisson and Laplace Equations	50
1.16	The Electric Dipole	51
1.17	Dipole Approximation	56
	Summary	59
	Problems	59
2	Conductors in Equilibrium	61
2.1	Conductors	63
2.2	Conductors in Equilibrium	64
2.3	Surface Charges on a Conductor	69
2.4	Hollow Conductors	72
2.5	Equilibrium in an Electrostatic Field	76
2.6	Electrostatic Capacitance	77
2.7	Calculating Capacitances	83
2.8	Combining Capacitors	85
2.9	Electrostatic Induction Coefficients	87

2.10	Electrostatic Shield	88
2.11	The Method of Images	91
	Summary	93
	Problems	94
3	Electrostatic Energy	97
3.1	Energy of a System of Point Charges	98
3.2	Energy of a Continuous Charge Distribution	99
3.3	Energy of a System of Conductors	101
3.4	Energy Stored in a Capacitor	102
3.5	Energy in the Electrostatic Field	103
3.6	The Energy of a Point Charge	107
	Summary	110
	Problems	110
4	Dielectrics	113
4.1	Dielectric Constant	114
4.2	Polarization of a Dielectric	116
4.3	Uniform Polarization	117
4.4	Non-uniform Polarization	119
4.5	Electrostatic Equations in a Dielectric	122
4.6	Linear and Isotropic Dielectrics	126
4.7	Electronic and Orientation Polarization	128
4.8	Electric Field in Cavities in Dielectrics	133
4.9	Electrostatic Energy in a Dielectric	138
	Summary	144
	Problems	145
5	Electric Currents	147
5.1	Current Intensity and Current Density	148
5.2	Electric Charge Conservation	154
5.3	Ohm's Law	156
5.4	The British Association Ohm	160
5.5	Surface Currents	162
5.6	Energy Balance	164
5.7	Generators	165
5.8	Slow Capacitor Discharge	167
5.9	Circuits in a Steady Regime	169
5.10	Superconductivity	171
	Summary	174
	Problems	174
6	Magnetostatics	177
6.1	Preliminary Observations	179
6.2	Magnetic Field	181
6.3	Hall Effect	184

6.4	Motion of Charges in a Magnetic Field	185
6.5	Galvanometer	195
6.6	The Magnetic Field of Steady Currents	196
6.7	Applications of Ampère's Law	203
6.8	The Vector Potential	209
6.9	The Vector Potential in Simple Cases	212
6.10	The Ampère-Laplace Law	216
6.11	Examples of Magnetic Field Calculations	217
6.12	Force Between Two Straight Steady Currents and the Ampere	219
6.13	The Magnetic Dipole	220
6.14	Charge Densities in a Current-Carrying Wire	225
6.15	Properties of the Laws Under Rotations and Inversions of the Axes	229
6.16	Relativity of Electric and Magnetic Forces	231
	Summary	237
	Problems	238
7	Electromagnetic Induction	241
7.1	The Flux Rule	242
7.2	Induced Electric Field. Faraday's Law	245
7.3	Exceptions to the Flux Rule	249
7.4	Betatron	250
7.5	Felici's Law	253
7.6	Energy Balance	255
7.7	Eddy Currents	257
7.8	Mutual Induction	260
7.9	Self-induction	262
7.10	Inductive Phenomena in Electric Circuits	266
7.11	Alternating Current Circuits	269
7.12	Complex Impedance	273
7.13	Energy Balance in a Circuit	279
	Summary	280
	Problems	280
8	Magnetic Energy	283
8.1	Energy of a Steady Current	284
8.2	Energy of a System of Steady Currents	285
8.3	Energy of a Dipole	286
8.4	Energy of the Magnetic Field	290
	Summary	294
	Problems	294
9	Magnetic Properties of Matter	297
9.1	Elementary Observations	298
9.2	Uniform Magnetization	300

9.3	Non-uniform Magnetization	303
9.4	Magnetic Field Equations in Matter	306
9.5	B and H Fields in Matter	309
9.6	Dia- and Para-magnetism. Microscopic Interpretation	314
9.7	Ferromagnetism	319
9.8	Uses of Ferromagnetism	323
9.9	Microscopic Interpretation of Ferromagnetism	327
9.10	Energy of Steady Currents in the Presence of Magnetic Materials	332
	Summary	335
	Problems	336
10	Maxwell Equations	339
10.1	Displacement Current	341
10.2	Electromagnetic Waves	345
10.3	The Maxwell Ratio of Units Experiment	350
10.4	Energy Density in the Electromagnetic Field	357
10.5	Energy Flux	360
10.6	Momentum Density in the Electromagnetic Field	363
10.7	Maxwell Equations in Matter	371
10.8	Discontinuities of E and B	377
10.9	The Electromagnetic Potentials	378
10.10	Covariance of Electromagnetism	382
10.11	Physical Meaning of the Electromagnetic Potentials	386
	Summary	394
	Problems	395
	Index	397

Symbols

Symbols for the Principal Quantities

a, a_s	Acceleration
α, α	Angular acceleration
ω	Angular frequency
I, \mathbf{L}	Angular momentum
Π	Canonical momentum
C	Capacitance
q, Q	Charge
ρ	Charge density
g	Conductivity
j	Current density
I	Current intensity
Γ	Curve
κ	Dielectric constant
q_e	Elementary charge
p	Electric dipole moment
D	Electric displacement
E	Electric field
Φ, Φ_E	Electric flux
χ_e	Electric susceptibility
\mathcal{E}	Electromotive force (emf)
U, U_E, U_m	Energy
w	Energy density (of the field)
S	Energy flux (of the field)
F	Force
v	Frequency
G	Gravitational field

Z	Impedance
L	Inductance
U_K	Kinetic energy
γ	Lorentz factor
\mathbf{H}	Magnetic auxiliary field
$\boldsymbol{\mu}$	Magnetic dipole moment
\mathbf{B}	Magnetic field
Φ, Φ_B	Magnetic flux
μ	Magnetic permeability (absolute)
κ	Magnetic permeability (relative)
\mathcal{R}	Magnetic reluctance
χ_m	Magnetic susceptibility
\mathbf{M}	Magnetization
m, M	Mass
$\langle x \rangle$	Mean value of x
\mathbf{p}	Momentum
\mathbf{g}	Momentum density (of the field)
M	Mutual inductance
G_N	Newton constant
T	Period
ϵ	Permittivity
θ, α	Plane angle
θ, ϕ	Polar angle
ρ, θ, ϕ	Polar coordinates (space)
\mathbf{P}	Polarization (density)
\mathbf{r}	Position vector
ϕ	Potential (electrostatic and scalar)
V	Potential difference
U_p	Potential energy
p, P	Pressure
r, R	Radius
Z	Reactance
μ	Reduced mass
R	Resistance
ρ	Resistivity
L	Self-inductance
Ω	Solid angle
S, Σ	Surface
σ	Surface charge density
\mathbf{k}, \mathbf{k}_s	Surface current density
t	Time
\mathbf{M}	Total moment
\mathbf{u}_v	Unit vector of \mathbf{v}

i, j, k	Unit vectors of the axes
n	Unit vector normal to a surface
ϵ_0	Vacuum permittivity
μ_0	Vacuum permeability
A	Vector potential
v, v	Velocity
<i>c</i>	Velocity of light (in a vacuum)
β	Velocity divided by light velocity
V	Volume
W	Work

Base Units in the SI

Quantity	Unit	Symbol
Length	metre/meter	m
Mass	kilogram	kg
Time	second	s
Current intensity	ampere	A
Thermodynamic temperature	kelvin	K
Amount of substance	mole	mol
Luminous intensity	candela	cd

Decimal Multiples and Submultiples of the Units

Factor	Prefix	Symbol	Factor	Prefix	Symbol
10^{24}	yotta	Y	10^{-1}	deci	d
10^{21}	zetta	Z	10^{-2}	centi	c
10^{18}	exa	E	10^{-3}	milli	m
10^{15}	peta	P	10^{-6}	micro	μ
10^{12}	tera	T	10^{-9}	nano	n
10^9	giga	G	10^{-12}	pico	p
10^6	mega	M	10^{-15}	femto	f
10^3	kilo	k	10^{-18}	atto	a
10^2	hecto	h	10^{-21}	zepto	z
10	deka	da	10^{-24}	yocto	y

Fundamental Constants

Quantity	Symbol	Value	Uncertainty
Light speed in a vacuum	c	299,792,458 m s ⁻¹	defined
Newton constant	G_N	$6.67308(31) \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$	47 ppm
Avogadro number	N_A	$6.022140857(74) \times 10^{23} \text{ mole}^{-1}$	12 ppb
Boltzman constant	k_B	$1.38064852(79) \times 10^{-23} \text{ JK}^{-1}$	570 ppb
Vacuum permittivity	$\epsilon_0 = 1/(c^2 \mu_0)$	$8.854187 817... \times 10^{-12} \text{ Fm}^{-1}$	defined
Vacuum permeability	$\mu_0 = 1/(c^2 \epsilon_0)$	$12.566370 614... \times 10^{-7} \text{ NA}^{-2}$	defined
Vacuum impedance	$Z = \mu_0 c$	376.730313461 ... Ω	defined
Elementary charge	q_e	$1.6021766208(98) ... \times 10^{-19} \text{ C}$	6.1 ppb
Unified atomic mass	$u=1\text{g}/N_A$	$1.660539040(20) \times 10^{-27} \text{ kg}$	12 ppb
Electron mass	m_e	$9.10938356(11) \times 10^{-31} \text{ kg}$	12 ppb
Proton mass	m_p	$1.672621898(21) \times 10^{-27} \text{ kg}$	12 ppb

Greek Alphabet

alpha	α	A	iota	ι	I	rho	ρ	P
beta	β	B	kappa	κ	K	sigma	σ, ς	Σ
gamma	γ	Γ	lambda	λ	Λ	tau	τ	T
delta	δ	Δ	mu	μ	M	upsilon	υ	Υ, Υ'
epsilon	ϵ	E	nu	v	N	phi	ϕ, φ	Φ
zeta	ζ	Z	xi	ξ	Ξ	chi	χ	X
eta	η	H	omicron	o	O	psi	ψ	Ψ
theta	θ, ϑ	Θ	pi	π	Π	omega	ω	Ω

Chapter 1

Electrostatic Field in a Vacuum

Abstract After defining the electric charge, we shall describe Coulomb’s experiment on the force between two point charges at rest in a vacuum and Millikan’s experiment on the elementary charge. We shall introduce the electric field and its potential, in time-independent conditions, and find and discuss the equations ruling these quantities, both in integral and differential form. We shall look at four important properties of electric charge, namely charge conservation, charge quantization, charge invariance and the equality of positive and negative elementary charges.

As recalled in the introduction, electromagnetic interaction is one of the four fundamental interactions known in nature. The other three are gravitational interaction, which was examined in Chap. 4 of the first volume of this course, and weak and strong nuclear interactions. The latter two are active at the microscopic level, are described by quantum mechanics and are beyond the scope of this course. Electromagnetic interaction is responsible for all macroscopic phenomena, gravitational examples aside. All fundamental interactions propagate in space through a “field” of force. For example, as discussed in the first volume, the sun generates a gravitational field in the space around it. An object, like the earth, at a point will then feel a force that is given by the field at that point times its own mass. Space is never empty. Even in a vacuum, where there is no matter, fields of force are always present, and energy and momentum with them, as we shall learn in subsequent chapters.

In this chapter, we study electric interaction in the simplest situation, namely between point charges at rest in a vacuum. After a few hints as to the charged constituents of matter, we shall describe the properties of the electric charge. We shall see that it is quantized, namely that the charge cannot be small at will. Contrastingly, an elementary charge exists, which is the charge of the electron and the proton. We shall also mention that electric charge can neither be created nor destroyed. Electric charge is always conserved. However, we must defer the quantitative description of charge conservation and of the governing equations to Sect. 5.2, after having introduced the concept of electric current. A third fundamental property is the invariance of the electric charge under transformations between inertial reference frames in relative motion. In other words, the electric charge of an object, for example, an electron or a nucleus, is independent of its

velocity, similar to its mass. We shall discuss the invariance of the charge in Sect. 1.9, after having gained the necessary knowledge.

In Sect. 1.2, we shall see that the force between two point charges at rest, as established by Coulomb, is very similar to the Newtonian force between two point-like (or spherical) masses. We shall also discover that the electric force obeys the superposition principle. The principle states that the force acting on the charge A due to the charges B and C together is equal to the vector sum of the force exerted on A by B when acting alone and that exerted by C when acting alone.

In Sect. 1.3, we shall introduce the concept of the electric field, under the particular conditions being discussed, namely produced by charges at rest. For this reason, we shall call it electrostatic, which simply means an electric field independent of time. An electric charge produces a “vector field” in the space around it independently of the presence of other charges. When a charged body, however, is present at a certain point, it will feel a force equal to the electric field at that point times its own charge. In Sect. 1.5, we shall see that the electrostatic field is conservative and hence define the electrostatic potential. In the subsequent two sections we shall show how to calculate an electric field and the potential for simple charge distributions.

Physics is an experimental science, based on the experimental method, which is the instrument for its progress. The history of physics is rich in fundamental and ingenious experiments. The study of a few allows appreciating how physics did and will progress. Two experiments are described in some detail in this chapter, the Coulomb experiment on the force between two charges in Sect. 1.2, as already mentioned, and the Millikan measurement of the elementary charge in Sect. 1.8.

After having defined the geometrical concept of solid angle in Sect. 1.10, we shall introduce the physical concept of the flux of an electric field in Sect. 1.11 and show its fundamental properties, which are stated as the Gauss law. In Sects. 1.12 and 1.13, we apply the Gauss law, calculating electric fields in symmetric geometries and finding the discontinuities of the field through a charged surface. The two important partial differential equations of Poisson and Laplace obeyed by the electrostatic potential are the objects of study in Sect. 1.15.

In the last two sections of the chapter, we shall study the simplest system beyond the point charge, namely the electric dipole, consisting of two equal and opposite point charges at a fixed distance. We shall consider both the field it produces and the torque and the force acting on a dipole in an external field.

1.1 Electric Charge

In the first years of VI Century BC, the Greek philosopher Thales of Miletus (Greece, c. 624–c 546 BC) reported that what we now know as an electric charge could be “produced” or, more accurately, accumulated by rubbing fur on several substances, amber in particular. Twenty-two centuries later, in 1600, the English scientist William Gilbert (UK, 1544–1603) published *De magnete* (for brief), in

which he not only dealt with magnetism, but also coined the word *electricus* from the Greek word *élektrikon*, meaning amber.

The electric charge is a fundamental property of matter. It is the source and receptor of one of the fundamental interactions of nature, electromagnetic interaction. As we shall learn in this book, electric, magnetic and chemical phenomena, and light itself, are all electromagnetic phenomena. Our civilization depends completely on electromagnetism. Just think about what our everyday life would be like if electric power were permanently cut.

There are two types of electric charges, called positive and negative. Charges of the opposite sign attract one another, charges of the same sign repel. Note that the adjectives ‘positive’ and ‘negative’ are just names; they do not mean that charges have the mathematical characteristics of positive and negative numbers.

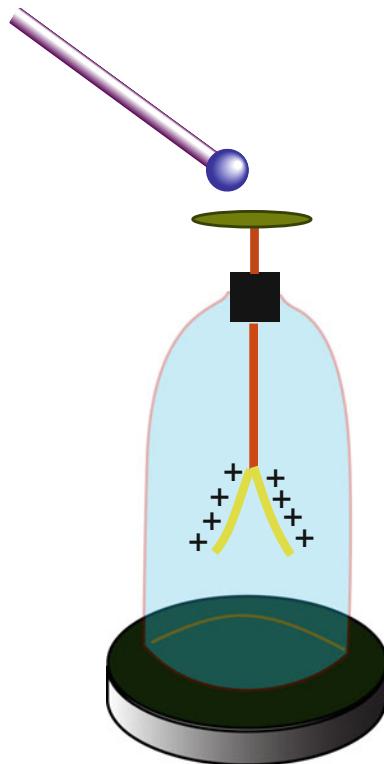
A simple device for a semi-quantitative determination of an electric charge is the electroscope, from the Greek word skopeo, to look. A common type, the gold-leaf electroscope, originally developed in 1787 by Abraham Bennet (UK, 1749–1799), is shown in Fig. 1.1. It consists of a vertical conductor bar, made of a metal, commonly brass. From the lower end of the bar hang two parallel thin flexible gold leaves, which are enclosed in a glass container to protect them from drafts of air. The container has a metallic base that is grounded for safety reasons. The upper end of the bar terminates in a metallic plate (or sphere), upon which the charge to be measured is deposited. Part of this charge spreads along the entire conductor, in particular, on the leaves. The leaves being charged with the same sign, they repel each other, opening into a “V” that gets wider as the charge gets larger. In this way, we have an evaluation, rather than a quantitative measurement, of the electric charge. Figure 1.1 also shows an *electrostatic spoon*, which is a small metallic sphere supported by an insulating handle. The spoon is used to take the charge from the point to be tested (for example, the surface of a conductor), to the electroscope, respectively touching the point of origin and then the electroscope’s plate with the sphere.

Instruments based on the same principle, called electrometers, have been developed to measure the electric charge more accurately.

Matter is made of very small particles, namely molecules that, in turn, are made of atoms. Atoms characterize the elements, molecules the chemical substances. The physical laws at the atomic and sub-atomic scales are quantistic and cannot be discussed at the level of this course. Consequently, we shall not enter into any detail, limiting the discussion to the basic elements, similar to what we did in Chap. 4 of the 2nd volume of this textbook.

Even if the atoms are the elementary objects in the chemical reactions, they have an internal structure. Atoms are composed of a central nucleus, which has a positive electric charge, and electrons, which are negative and form a “cloud” around the nucleus. Atoms are electrically neutral; the binding force is electromagnetic. The atomic nucleus has an internal structure as well; it is made of protons and neutrons. The force keeping the nucleus together is called the strong nuclear force. Protons and neutrons are also composite objects; they are made of quarks, bound by the so-called color force, for which “color” is a funny name given to it by physicists,

Fig. 1.1 Gold-leaf electroscope and electrostatic spoon



despite having nothing to do with color as we know it. The strong nuclear force, to be precise, is a consequence of the color force.

The geometrical dimensions of the atoms, different from one atomic species to another, are the dimensions of this negative “cloud”. The order of magnitude is a tenth of a nanometer, or 10^{-10} m. The nuclear diameters are four orders of magnitude smaller, between 1 fm and 10 fm (10^{-15} – 10^{-14} m). If we were to magnify a nucleus to the size of the dot on an “i” on this page, the atom would be a few meters in size.

The number of electrons (symbol e), called Z , which characterizes the element, varies from 1 for hydrogen to 92 for uranium in the Mendeleev table, which is named after Dimitri Mendeleev (Russia, 1834–1907). The electrons inside atoms behave according to quantum, rather than classical, laws. In particular, electrons do not have well-defined trajectories; we cannot properly speak of electron orbits around the nucleus (even this is still found in many books). Atomic electrons move very fast compared to macroscopic objects; their speeds are on the order of 10^4 m/s, which, however, are much smaller than the speed of light. The characteristic times of the electrons’ motion are much smaller than the resolving times of our instruments and, consequently, we observe an average configuration of the atom. We see the electron charge as continuously distributed in a region around a nucleus with

larger density where the probability of finding one electron is greater, and more sparsely where the probability is smaller. We can then think of a cloud of charge even if there is only one electron, as in the case of hydrogen.

The atomic nucleus is made of protons (p) that are positive and neutrons (n) that are neutral. The electric charge aside, protons and neutrons are very similar and are collectively called nucleons. For every element, the number of protons is equal to the number of electrons. Protons and electrons have equal and opposite charges; atoms, as we said, are globally neutral.

The proton and electron electric charge is the smallest charge existing *free* in nature, and is called the *elementary charge*. As a matter of fact, quarks have smaller charges. Nucleons contain two types of quark, called *up* (u) and *down* (d). Their charges are $2/3$ and $-1/3$ of the elementary charge, respectively. The proton contains $2 u$ and $1 d$, the neutron $1 u$ and $2 d$. However, quarks are never free; they live inside the nucleons and other particles of the same category. The charges of all the other objects are integer multiples of the elementary charge. One might consider adopting the elementary charge as the unit, but this is not convenient, because enormous numbers would represent all the usual charges. The unit of electric charge in the SI is the *coulomb* (C), which we shall soon define. The value of the elementary charge is

$$q_e = 1.6021766208(98) \times 10^{-19} \text{ C}. \quad (1.1)$$

where the two digits in parentheses are the uncertainty of the last two digits of the reported value. We shall describe, in Sect. 1.8, the measurement by Millikan of this fundamental quantity.

The value of the elementary charge is commonly used to define an energy unit that is useful at the atomic and molecular scales, which is called the electronvolt. The electronvolt is the kinetic energy gained by an electron falling under the potential difference of one volt. As such, its numerical value in joule is equal to the elementary charge, namely, in a round figure,

$$1 \text{ eV} \cong 1.60 \times 10^{-19} \text{ J}. \quad (1.2)$$

The presence of neutrons in the nucleus is necessary to guarantee its stability. Inside the nucleus, the repulsive electric force between protons tends to destroy it. The nuclear force is, however, attractive and, under the same conditions, has the same intensity between protons, between neutrons and between a proton and a neutron. The balance between electric and nuclear forces is realized when the number of neutrons is somewhat larger than the number of protons. The neutron excess increases with increasing nuclear size. The number of neutrons is indicated with N , the total number of nucleons (protons plus neutrons) with A ($A = N + Z$), which is called the *atomic number*. For a given atomic species (namely a given Z), more than one nuclear species may exist, with different values of N and consequently of A . All of them have the same chemical properties and are lodged in the same box of the Mendeleev table. For this reason, they are called *isotopes* (=‘same

place' in Greek). The percentages of the different stable isotopes of an element are fixed in nature.

For example, hydrogen has two stable isotopes: ^1H (the superscript is A), the nucleus of which is simply the proton, and ^2H , the deuteron, the nucleus of which is made of a proton and a neutron. A third isotope, tritium ^3H , exists but is unstable, having a half-life of 12.32 years. It is continuously produced by cosmic ray collisions in the atmosphere. The second element is helium, which has two stable isotopes, ^3He ($2p, 1n$) and ^4He ($2p, 2n$), and so on.

Let us now see the values of the masses. The electron mass is

$$m_e = 9.10938356(11) \times 10^{-31} \text{ kg}. \quad (1.3)$$

The proton is 1836 times larger, namely

$$m_p = 1.672621898(21) \times 10^{-27} \text{ kg}. \quad (1.4)$$

The neutron has a mass almost equal to, but a bit larger than, the proton, namely

$$m_n = 1.674927471(21) \times 10^{-27} \text{ kg}, \quad (1.5)$$

From the above values, we see that the largest fraction of the atom mass, and with them of the mass of matter, is concentrated in the nucleus. The electrons' contribution is only a few parts in ten thousand. One might think that the atomic masses of the elements are integer multiples of the proton mass. This is so only in a rough approximation, for three reasons. First, several elements are a mixture of different isotopes with different values of A , in some proportions; second, the proton and neutron masses are almost, but not exactly, equal; third, the mass of the nucleus is not equal to the sum of the masses of its nucleons; it is smaller than that due to the binding energy, as discussed in Chap. 6 of the 1st volume.

The masses of the u and d quarks are about two per mille and four per mille of the nucleon mass, respectively. This is really surprising. From where does the largest fraction of the mass of the nucleons, and consequently of the nuclei, of the atoms, of matter in general, come? The answer is in the very peculiar behavior of the (quantum) color force. On one side, it increases with the distance so much that quarks cannot be taken apart; on the other, its binding energy is *positive* and very large. Consequently, instead of a mass defect, such as in atoms and nuclei, in the nucleons, there is a mass *excess*. Namely, the mass of the nucleon is much larger than the sum of the masses of its components. This excess is the largest fraction of the mass of matter, we included.

As far as we know, electrons and quarks do not have an internal structure and are point-like. Namely, their sizes, if any, are smaller than the experimental resolution, which is presently on the order of the attometer (10^{-18} m). The nucleons have a radius smaller than, but comparable to, nuclei, of about 1 fm, and, as we already said, are composed of quarks. In our discussion of dielectrics, conductors and

magnetic materials, we shall not need to know more details of the internal structure of nuclei.

In conclusion, matter is made of an enormous number of very small charged elementary constituents, with electric charges of both signs, which are so intimately and precisely mixed that their effects, which are enormous inside the atoms, almost disappear outside them. We shall see that the forces between charges are similar to the gravitational force, but far more intense. The fact that their effects are almost unnoticeable at the macroscopic level is due to their perfect cancellation at the atomic level, resulting in the global neutrality of matter.

QUESTION Q 1.1. What is the charge of a mole of electrons? □

Electric charge conservation. A fundamental property of electric charge is its conservation. The creation or destruction of electric charge is impossible, in the sense that the total charge, the algebraic sum of positive and negative charges, cannot vary. It is true that, for example, an electron can “annihilate” with its antiparticle, the positron, which has an equal and opposite charge in the final state of two photons, which are neutral. But both the initial and final charges are globally zero. Similarly, the process in which a photon “materializes” in an electron positron pair is possible, because the total charge is conserved. Another example is given by the neutron, which, when free outside a nucleus, is unstable. As a matter of fact, neutrons decay, by the weak nuclear force into a proton, an electron and an antineutrino. Their charges are $+q_e$, $-q_e$ and 0, respectively. The total charge is conserved.

Notice than when one talks, somewhat colloquially, of “producing” charge, for example, by rubbing fur on amber, as we did in the first lines of this section, one always means separating some negative charges (that remain on the fur) from positive ones (that remain on the amber).

1.2 Coulomb's Law

In 1785, using the torsion balance he had invented, Charles Augustin de Coulomb (France, 1736–1806) demonstrated that two like charges repel, and two unlike charges attract, each other with a force that varies according to the inverse square of their distance. Before describing this famous experiment, we must state the circumstances under which two charges are equal and establish what multiple and submultiple charges are. We say that two charges are equal, in value and sign, when they separately exert equal forces in magnitude and direction on a third charge at the same distance from both. The two charges are equal in magnitude and opposite in sign, if, under the same conditions, they exert forces equal and opposite. Before accepting the definition, we need to verify experimentally that what we have defined as equality indeed behaves as equality does. We must ask: is the transitive property satisfied? Namely, if, with the definition we have given, charge A is equal to charge B and charge B is equal to charge C , are charges A and C equal? The experimental answer is yes. Notice that this conclusion cannot be reached through pure logic; it needs to be experimentally verified.

As for the multiples, a charge n times a given charge is obtained putting together n charges equal to the given one. We obtain the submultiples by dividing the charge under symmetric conditions. We give the charge to be divided to a small conducting, and insulated, sphere. We then touch it with an identical sphere that has not been charged, taking care that the surrounding bodies respect the symmetry between the two spheres. The initial charge divides into two equal parts on the two spheres. We finally verify experimentally, as in the case of equality, that our definitions are coherent.

In his first memoire to the French Royal Academy of Sciences in 1785, Coulomb starts by explaining

how to construct and use an electric balance [*he means a torsion balance*] based on the property of the metal wires of having a reaction torsion force proportional to the torsion angle

For a discussion of the torsion balance, see Sect. 8.9 in the 1st volume. Figure 1.2a, b, taken from the memoire, show the apparatus. A thin metal (Ag) fiber is attached to the suspension head on the top of the apparatus. The lower end of the fiber, at the bottom of a column and in a glass cylinder, is attached at the middle point of the balance crossbar (Fig. 1.2c). The equilibrium angle of the balance, in absence of applied torque, can be adjusted acting on the suspension head and read out from the position of an index moving on a scale on the head. The torque exerted by the wire when the crossbar is rotated out of equilibrium is proportional to the rotation angle, as in the above sentence by Coulomb. The rotation angle is measured on a scale placed around the glass cylinder (Fig. 1.2a). This gives the torque, having calibrated the balance in advance. A small conducting sphere is fixed at one end of the crossbar, a counterweight at the other end. The sphere is charged at the beginning of the experiments. A second conducting sphere can be placed near the suspended one, through an opening at the top of the glass cylinder, by means of the small bar shown in Fig. 1.2b. In each experiment, it is charged outside and then introduced into the apparatus.

Coulomb started charging the spheres with the same sign. In his first measurement, he set the unperturbed equilibrium position of the balance at a certain angle θ_{10} . The equilibrium between the torque of the electric force and the elastic torque of the fiber was reached when the centers of the spheres were at the distance, say, d_{1f} , and the angle of the balance was θ_{1f} . The torque (and the force) was proportional to $\theta_{1f} - \theta_{10}$. In the second measurement, he changed the unperturbed position to, say, θ_{20} so that the equilibrium distance between the spheres would be one half of the first measurement, namely $d_{2f} = d_{1f}/2$. The corresponding angle was θ_{2f} . Coulomb found that the torque, proportional to $\theta_{2f} - \theta_{20}$, had quadrupled. He halved the final distance once more, finding the torque quadrupled again. He thus concluded:

It follows therefore from these three tests, that the repulsive action that the two balls electrified of electricity of the same type exert on one another varies in inverse proportion to the square of the distance (*between centers*).

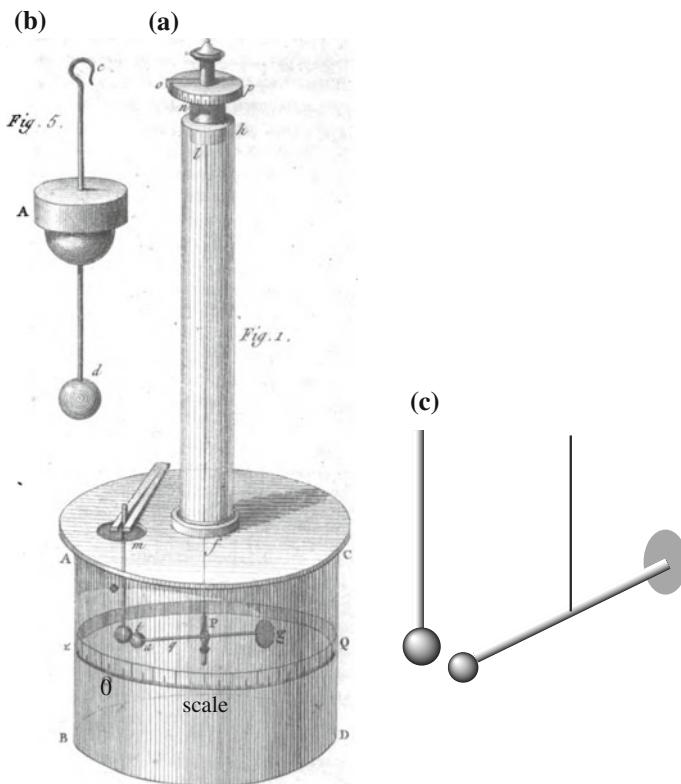


Fig. 1.2 Coulomb's torsion balance experiment. **a** the apparatus, **b** detail showing the fixed sphere; both are from the memoir, **c** torsion balance detail, showing the crossbar, charged spheres and counterweight

The attractive force is more difficult to measure, because if the product of the charges is larger than a certain angle-dependent value, the electric torque cannot be equilibrated by the elastic torque of the wire. The balls approach, touch each other and discharge. Coulomb found, however, that an allowed operational range exists, in which he performed his measurements. He published the results in a second memoir in the same year, in which he stated

Finally, comparing the different experiments, I concluded that the attractive force between electrified balls, one of electricity that I shall call *positive*, the other of that I shall call *negative*, is inversely proportional to the square of the distance, as already found for the repulsive force.

Other experiments, halving the charge of the sphere of Fig. 1.2b, showed that the force is proportional to the product of the charges. The conclusion is Coulomb's law

$$F = k \frac{q_1 q_2}{r^2}. \quad (1.6)$$

The force is also called Coulomb's force or an electrostatic force. Let us make a few observations. The first one is that Coulomb's law holds only for the force between two charges at rest (the term "electrostatic" recalls that). It does not hold if the charges are in motion, as we shall learn subsequently. In addition, the law was established for the force between point charges. However, Coulomb's force depends on the distance as the gravitational force. Consequently, the force between two spherical distributions of charges, as in the case of the Newton force between masses, is equal to the force between those charges concentrated in the centers of the spheres. Coulomb's law is rigorously valid in a vacuum. We shall see the effects of a medium in Chap. 3. We anticipate here that if the medium is air, as in Coulomb's experiment, the effects are extremely small. Finally, we observe that Coulomb's experiment had a limited sensitivity. The inverse square distance dependence of the force is rigorously established by observing the electrostatic shielding effects, which exist only for the inverse square behavior. These effects, which we shall study in Sect. 2.10, are analogous to the property of the Newton force of being zero inside a spherical mass shell.

The value of the proportionality constant k depends on the measurement unit chosen for the electric charge. In the SI, the base quantity is the current intensity, which is the electric charge that goes through a section of a conductor in a second. This is called an *ampere*, after André-Marie Ampère (France, 1775–1836), and has the symbol A. The definition of the ampere is based on the magnetic effects of the current, and we must defer study of it to Sect. 6.12. The unit of charge is the charge passing in one second in a circuit traversed by a stationary current of 1 A. It is called a *coulomb* and its symbol is C, namely $1 \text{ C} = 1 \text{ A} \times 1 \text{ s}$.

In the SI, the proportionality constant k is defined in terms of another constant, ϵ_0 , called the *electric permittivity* of the free space or *vacuum permittivity*, as

$$k = \frac{1}{4\pi\epsilon_0}. \quad (1.7)$$

The electric permittivity of the free space is a fundamental constant of nature. As we shall learn in this course, in the SI system, it is linked to a fundamental invariant quantity, the speed of light in a vacuum. The value of the latter, as known from the study of mechanics, is given by the definitions of the second and of the meter. The consequence is that the value of ϵ_0 is also fixed by those definitions. As such, it is said to be "exact", having moved the experimental uncertainties in the definitions. Its value is

$$\epsilon_0 = 8.854187817 \times 10^{-12} \text{ N}^{-1} \text{ m}^{-2} \text{ C}^2. \quad (1.8)$$

We shall meet ϵ_0 continuously in the following calculations and we should remember the first digits of its value by heart. This is more easily done using the

measurement unit of the electrostatic capacitance, which is the farad (F). We shall define the farad in Sect. 2.6. Here, we need only know that the dimensions of ϵ_0 are capacitance per unit length, namely the unit is 1 F/m. We write the first digits of Eq. (1.8) in the easy to remember form of

$$\epsilon_0 = 8.8 \text{ pF/m.} \quad (1.9)$$

Equation (1.6) expresses the magnitude of the force. To obtain its vector expression, let us consider the point charge q_1 at rest in an inertial frame at the position vector \mathbf{r}_1 and the point charge q_2 at rest at the position vector \mathbf{r}_2 and let $\mathbf{r}_{12} = \mathbf{r}_2 - \mathbf{r}_1$ be the vector from the first to the second charge and \mathbf{u}_{12} its unit vector, as shown in Fig. 1.3. The electrostatic force exerted by q_1 on q_2 is

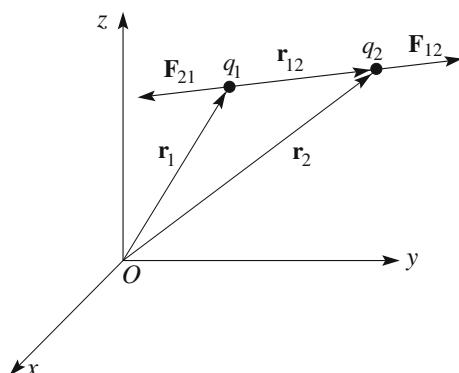
$$\mathbf{F}_{12} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r_{12}^2} \frac{\mathbf{r}_{12}}{r_{12}} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r_{12}^2} \mathbf{u}_{12}. \quad (1.10)$$

The expression shows, in particular, that if the charges have the same sign, \mathbf{F}_{12} and \mathbf{r}_{12} have the same direction, namely the force is repulsive, while if the charges have opposite signs, \mathbf{F}_{12} and \mathbf{r}_{12} have opposite directions, namely the force is attractive.

Let \mathbf{F}_{21} be the electrostatic force exerted by q_2 on q_1 . The question arises as to whether \mathbf{F}_{21} is equal and opposite of \mathbf{F}_{12} and on the same application line. In other words, does the action-reaction law hold for these forces? The answer must be given by experiments, and is positive. Notice that this is not at all a priorly obvious. A counter example is given by the electric charges in motion. The forces they exchange, as we shall see in Sect. 10.6, do not obey the second Newton law.

Consider now a point charge q_0 at rest at \mathbf{r}_0 and a number, say n , of other charges, which are point like and at rest too, say q_1 at \mathbf{r}_1 , q_2 at \mathbf{r}_2 , q_3 at \mathbf{r}_3 ... q_n at \mathbf{r}_n . It is experimentally found that the superposition principle holds for the electrostatic forces, namely that the total force is the sum of the forces that each charge would exert if acting alone. If $\mathbf{r}_{i0} = \mathbf{r}_i - \mathbf{r}_0$ is the vector from the generic charge q_i to q_0 and \mathbf{u}_{i0} is its unit vector, the force on q_0 is

Fig. 1.3 Electrostatic forces between two point charges at rest in a vacuum



$$\mathbf{F} = \frac{q_0}{4\pi\epsilon_0} \sum_{i=1}^n \frac{q_i}{r_{i0}^2} \mathbf{u}_{i0}, \quad (1.11)$$

Coulomb's force between two point charges q_1 and q_2 at rest in a vacuum is very similar to the Newton force between two point-like masses m_1 and m_2 , which is, we will recall,

$$\mathbf{F}_{12}^N = -G_N \frac{m_1 m_2}{r_{12}^2} \mathbf{u}_{12}. \quad (1.12)$$

where the minus sign tells us that the force is attractive in any case. This is an important difference between electrostatic and gravitational forces. The second important difference is their magnitude. We recall that the value of the Newton constant is $G_N = 6.67 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$. Take, for example, a proton and an electron in a hydrogen atom. They act on one another with the electrostatic and gravitational forces. We can compare their ratio at any distance, said ratio being independent of distance because the two forces vary in the same way. We have

$$\frac{F_{\text{electrost}}(ep)}{F_{\text{gravitaz}}(ep)} = \frac{q_e^2}{4\pi\epsilon_0 G_N m_e m_p} \cong 10^{39}.$$

This number is very huge indeed. Consider two heavenly bodies, the earth and the moon, for example. The protons and electrons of one of them act on the protons and electrons of the other, both electrically and gravitationally in the intensity ratio we have just seen. But we do not see any resulting electric force; we see only the gravitational one. This is because positive and negative charges are intimately and accurately mixed in matter and because the proton and electron charges are, in absolute value, exactly equal.

As an example, suppose the electron charge to be a bit larger than the proton charge, say of a part in a billion, i.e., $|q_e| = 1.000000001 q_p$. What is the electrostatic attraction between two iron spheres each of 1 kg mass at the distance of 1 m? A Fe atom has 26 electrons, 26 protons and 29 neutrons. Its molar mass is, consequently, about 55 g. Each sphere contains $(1000/55) \times 6.02 \times 10^{23} = 1.1 \times 10^{25}$ atoms, and consequently, 2.8×10^{26} electrons and as many protons. In the above hypothesis, the charge of each sphere would be $q = 2.8 \times 10^{26} \times 1.6 \times 10^{-19} \times 10^{-9} = 4.6 \times 10^{-2} \text{ C}$. The electrostatic attraction would be $(4.6 \times 10^{-2})^2 / (4\pi \times 8.8 \times 10^{-12}) = 2.4 \times 10^7 \text{ N}$, which is about the weight of 2400 t. We shall come back to charge equality in Sect. 1.9.

QUESTION Q 1.2. Evaluate the ratio between electrostatic and gravitational forces between two electrons at 1 mm, and then 1 m, distance. Do the same for two protons. \square

1.3 The Electrostatic Field

The concept of a field of forces is central in physics. In the first volume of this course, we studied the gravitational field, which is the field of the Newton force. Similarly, the field of Coulomb's force is the *electrostatic field*, which we shall now define. More generally, in this book, we shall study the properties of the electric field, which is, in general, a function of position and time. When it does not depend on time, namely is constant, it is called electrostatic.

The electrostatic field is produced by electric charges. Let us consider the simplest case, namely a single, point charge q_1 at rest in the position \mathbf{r}_1 , which we shall call the source of the field. Its field is a mean for describing its action on other charges. Let q_0 be such a charge, \mathbf{r}_0 its position, $\mathbf{r}_{10} = \mathbf{r}_1 - \mathbf{r}_0$, the vector from q_1 to q_0 and \mathbf{u}_{10} its unit vector. The electrostatic force on q_0 is

$$\mathbf{F}_{10} = q_0 \frac{1}{4\pi\epsilon_0} \frac{q_1}{r_{10}^2} \mathbf{u}_{10}, \quad (1.13)$$

which is proportional to q_0 , namely the charge that "feels" the force and that we shall call the explorer charge, and to a vector term, which depends on q_1 and its position

$$\mathbf{E}(\mathbf{r}_0) = \frac{1}{4\pi\epsilon_0} \frac{q_1}{r_{10}^2} \mathbf{u}_{10}, \quad (1.14)$$

This is the electric field produced by the source q_1 . The force acting on q_0 is

$$\mathbf{F}_{10} = q_0 \mathbf{E}(\mathbf{r}_0), \quad (1.15)$$

The just made argument did not introduce anything new in addition to Coulomb's law; it has just changed the point of view. Indeed, we have introduced an asymmetry in the way of thinking about the two charges. We have considered one of them, namely q_1 , as giving origin to an entity, the electric field, at all the space points around q_1 , and we have called it the source of the field for this reason. The next step of the argument is that, if we place another charge q_0 at a generic point in space, this charge, which we call the receptor, feels a force equal to its own value times the field at that point, which is independent of q_0 . The implication of this way of thinking is that the field also exists when q_0 is not there. To measure the field at a point, we must put an exploring charge at that point, measure the force on it and divide this force by the charge we have used.

The concept of field becomes more useful when we deal with sources more complex than a single charge. Indeed, Eq. (1.11) tells us that the force felt by the charge q_0 in the presence of n other point charges is also the product of q_0 and of a vector independent of q_0 , which we shall again call the electric field, namely

$$\mathbf{F} = q_0 \mathbf{E}(\mathbf{r}_0) \quad (1.16)$$

with

$$\mathbf{E}(\mathbf{r}_0) = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^n \frac{q_i}{r_{i0}^2} \mathbf{u}_{i0}. \quad (1.17)$$

We recognize the right-hand side immediately as being the sum of terms as in Eq. (1.14), one for each of the n charge sources of the field. These are the fields that each of them would produce if alone. Clearly, the superposition principle also holds for the electric field.

Equation (1.17) is a vector equation. Let us write one of its Cartesian components (the other ones are obviously similar) explicitly in terms of the coordinates of $\mathbf{r}_i = (x_i, y_i, z_i)$ and $\mathbf{r}_0 = (x_0, y_0, z_0)$, remembering that $\mathbf{u}_{i0} = \mathbf{r}_{i0}/r_{i0}$. We have

$$E_x(x_0, y_0, z_0) = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^n \frac{q_i(x_0 - x_i)}{\left[(x_0 - x_i)^2 + (y_0 - y_i)^2 + (z_0 - z_i)^2\right]^{3/2}}. \quad (1.18)$$

This equation, and those analogous for the other components, must be used to calculate the field of any system of point charges at rest of known values and in known positions.

It is often convenient to think of a continuous distribution of charges, ignoring their discrete nature. This is certainly the case for macroscopic objects that generally contain an enormous number of microscopic charges. If Δq is the charge in the small volume ΔV at (x, y, z) , the charge density

$$\rho(x, y, z) = \lim_{\Delta V \rightarrow 0} \frac{\Delta q}{\Delta V} = \frac{dq}{dV} \quad (1.19)$$

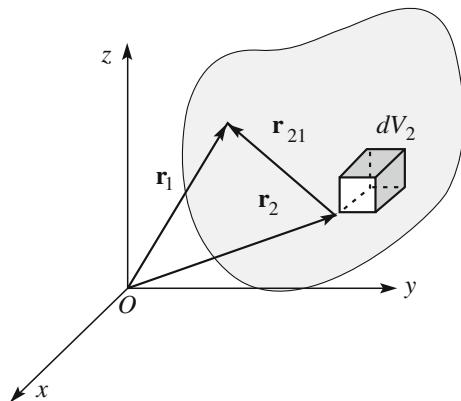
where the limit is meant for volumes very small relative to the macroscopic dimensions but still big enough to contain a large number of molecules. This is completely similar to the definition of mass density. The physical dimensions of charge density are a charge per unit volume, namely $C m^{-3}$.

Consider a generic continuous charge distribution, represented by a grey area in Fig. 1.4. The electric field at the point $\mathbf{r}_1 = (x_1, y_1, z_1)$ is the sum of the contributions of the charge elements $dq = \rho dV_2 = \rho dx_2 dy_2 dz_2$ located at $\mathbf{r}_2 = (x_2, y_2, z_2)$, namely

$$\mathbf{E}(\mathbf{r}_1) = \frac{1}{4\pi\epsilon_0} \int_V \frac{\rho(\mathbf{r}_2) \mathbf{u}_{21}}{r_{21}^2} dV_2. \quad (1.20)$$

where V is the volume of the charge distribution source of the field.

Fig. 1.4 Calculating the electric field of a continuous charge distribution



An electric charge may be distributed along surfaces or lines, rather than in volumes. The electric field generated by such distributions is given by an expression similar to Eq. (1.20), with integrals on those surfaces or lines, rather than a volume. In these cases, we deal with surface charge density, namely the charge per unit surface, or the linear charge density, namely the charge per unit length.

The expressions we have found allow us to calculate the static electric field of a given charge distribution at every point in space.

Let us now look more carefully at the operational definition of the electric field. We want to measure the electric field at a generic point \mathbf{r} , independent of whether or not we know its sources. We must place a known charge q_0 at rest in \mathbf{r} and measure the force $\mathbf{F}(\mathbf{r})$ on it. The field is then $\mathbf{E}(\mathbf{r}) = \mathbf{F}(\mathbf{r})/q_0$. To be precise, however, we should consider that when we place the exploring charge in \mathbf{r} , that charge necessarily exerts forces on the charge sources of the field. These forces will induce changes in the field we are measuring. The smaller this unwilled effect is, the smaller the exploring charge. Let $\Delta\mathbf{F}(\mathbf{r})$ be the force on the exploring charge Δq at rest in \mathbf{r} . Then, the electric field in \mathbf{r} is defined as

$$\mathbf{E}(\mathbf{r}) = \lim_{\Delta q \rightarrow 0} \frac{\Delta\mathbf{F}(\mathbf{r})}{\Delta q} \quad (1.21)$$

In this way, we can know, in principle, the electric field at any point we are interested in with a series of measurements, even ignoring the sources of the field. We observe that, in practice, the exploring charge Δq has to be lodged in a material body. We can think of a small metallic sphere to fix ideas. The force we measure is, consequently, the effect of the field averaged on the dimensions of the sphere. These dimensions may be small, but cannot be less than some reasonable limit, which is certainly enormous in comparison to the molecular dimensions. In addition, the measurement operation requires some time; it cannot be instantaneous. This time is

also enormous in comparison with the characteristic molecular times. The field at point \mathbf{r} in the instant t , which is determined with our operational definition, must be thought of as an average over a volume around \mathbf{r} , which is very small on macroscopic dimensions but very large on the nanometric scale, and on a time interval around t , small on the macroscopic scale but long on the molecular phenomena scale (femtosecond).

From the conceptual point of view, the field concept eliminates the action at a distance. Let us consider the simple case of two point charges q_1 and q_0 at rest. The physical fact we observe is a force on q_0 and an equal and opposite force on q_1 . Each of them is an action at a distance, for example, of q_1 on q_0 . With the field concept, we think that q_1 creates an electric field all around, even when there is no q_0 to sense it. When q_0 is present, the field manifests itself as a force at the point where q_0 is. The effect is now local. The two descriptions are completely equivalent as long as we are in static conditions. In other words, in electrostatics, it is impossible to answer the question: does the electric field really exist when there is no exploring charge to sense it? The answer to this question, however, exists and is a yes under general dynamic conditions. We shall see that when the field changes in time, there are physical phenomena that show the presence of the field without the need for any explorer charge. Empty space in which an electric field is present does not contain matter, but it does contain something physical, something that we measure. It contains energy, linear and angular momenta. We shall also see that the description of the interactions between charges is much simpler in terms of fields (electric and magnetic) than in terms of forces.

The electric field has physical dimensions of a force per unit charge. Its measurement unit is one newton per coulomb (N/C) or, equivalently and more often used, the volt per meter (V/m).

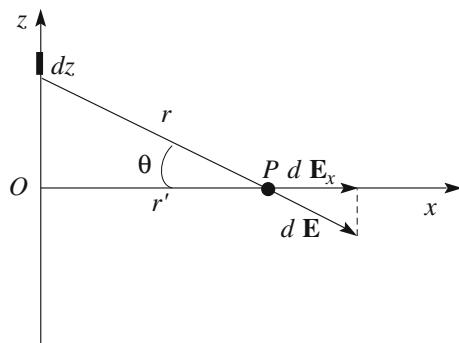
1.4 Calculating Electric Fields

In this section, we shall give two examples of calculation of an electric field produced by geometrically simple charge distributions. We should mention that computer applications are available to calculate numerically the field of any charge distribution.

Linear charge distribution.

Consider a charge distribution along an infinite straight line with uniform linear density (namely charge per unit length) λ . We calculate the field at the generic point P at a distance r' from the line. Let us take the z -axis on the line, as shown in Fig. 1.5. Given the symmetry of the problem, the field in P can depend only on r' and we can take the x -axis through P without losing the generality of the argument. Let us start by expressing the contribution to the field of the charge element between z and $z + dz$ (see Fig. 1.5), which is

Fig. 1.5 Infinitesimal contribution to the electric field of an element of a linear uniform charge distribution



$$dE = \frac{1}{4\pi\epsilon_0} \frac{\lambda dz}{r^2}. \quad (1.22)$$

The symmetry of the problem tells us that the total field is necessarily perpendicular to the line, namely in the x direction. As a consequence, we shall sum (integrate) the elementary x -components, namely

$$dE_x = \frac{1}{4\pi\epsilon_0} \frac{\lambda dz}{r^2} \cos \theta. \quad (1.23)$$

We take θ as the integration variable between $-\pi/2$ and $+\pi/2$. Expressed in this variable, the other quantities are $r = r'/\cos \theta$, $z = r' \tan \theta$ and $dz = (r'/\cos^2 \theta)d\theta$ and the integral of Eq. (1.23) is

$$E_x = \int_{-\pi/2}^{+\pi/2} dE_x = \frac{\lambda}{4\pi\epsilon_0 r'} \int_{-\pi/2}^{+\pi/2} \cos \theta d\theta = \frac{\lambda}{2\pi\epsilon_0 r'}. \quad (1.24)$$

Observe that the field decreases as the inverse of the distance from the wire rather than with the inverse squared. This is because, at any given distance, the main contribution comes from a length of wire of the same order of the distance itself. Moving away, the contribution of each element decreases as the inverse distance squared, but the number of those that substantially contribute increases in proportion to the distance.

Planar charge distribution

Consider an indefinite plane of uniform charge density (charge per unit surface) σ . Let P be a point at a distance r' from the plane. We take a reference frame with the y and z axes in the plane and the x -axis through P . This choice does not affect the generality of the argument, because the symmetry of the problem implies that the field in P depends only on its distance from the plane.

The symmetry of the problem also implies that the total field is perpendicular to the plane. As in the previous example, we can only consider the sum of the x -components of the charge elements. We lump together all the elements whose field has the same x -component. These are the elements of the annuluses between ρ and $\rho + d\rho$, as shown in Fig. 1.6.

The contribution of the infinitesimal annulus, whose area is $2\pi\rho d\rho$, is

$$dE_x = \frac{\sigma}{4\pi\epsilon_0} 2\pi\rho d\rho \frac{\cos\theta}{r^2}. \quad (1.25)$$

We now express it all in terms of r . First, observe that, given that $r'^2 + \rho^2 = r^2$ and r' being fixed, we have $\rho d\rho = rdr$. In addition, it is $r' = r \cos\theta$. Substituting in Eq. (1.25) and integrating r from its minimum value, which is r' , to infinity, we have

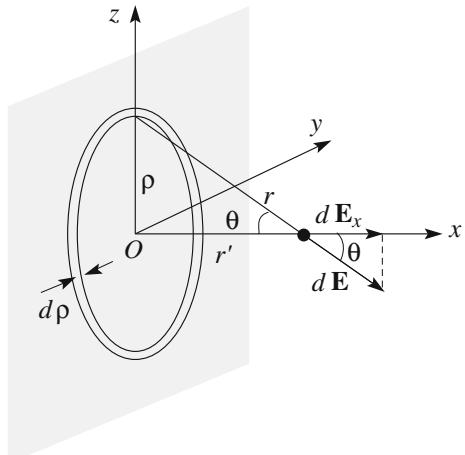
$$E_x = \frac{\sigma}{2\epsilon_0} r' \int_{r'}^{\infty} \frac{dr}{r^2} = \frac{\sigma}{2\epsilon_0} \quad (x > 0). \quad (1.26)$$

We have specified that the solution holds for $x > 0$ because the field is directed away from the plane when $\sigma > 0$, and towards the plane when $\sigma < 0$. Consequently, E_x has the same sign of σ for $x > 0$, as in Eq. (1.26), and the opposite sign for $x < 0$. Namely, it is

$$E_x = -\frac{\sigma}{2\epsilon_0} \quad (x < 0). \quad (1.27)$$

We observe that the field, which is normal to the charged surface, has a discontinuity across the surface equal to σ/ϵ_0 . In Sect. 1.13, we shall see that this property holds for charged surfaces of any shape.

Fig. 1.6 Infinitesimal contribution to the electric field of an element of a planar uniform charge distribution



We also notice that the magnitude of the field is independent of the distance. This is so because when the distance increases, the number of substantially contributing charge elements increases as the distance squared. The case of an infinite plane is, however, clearly unrealistic. Nonetheless, Eqs. (1.26) and (1.27) are good approximations at points sufficiently close to the plane.

1.5 Electrostatic Potential

We shall now show that the electrostatic force produced by any set of charges at rest on a generic charge q is conservative. The corresponding field is said to be conservative as well. We first demonstrate the statement for the field produced by a single point charge q_0 and then generalize the result to any number of charges using the superposition principle.

When dealing with a single charge q_0 , it is convenient to choose a reference frame with the origin O in its position, as in Fig. 1.7. The force on the charge q in \mathbf{r} (unit vector \mathbf{u}_r) is

$$\mathbf{F}(\mathbf{r}) = q \frac{q_0}{4\pi\epsilon_0 r^2} \mathbf{u}_r. \quad (1.28)$$

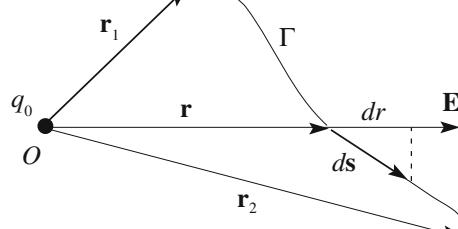
Let us calculate the work needed to move the charge q from the initial position vector \mathbf{r}_1 to \mathbf{r}_2 , along a certain trajectory Γ . We shall show that the work is independent of the trajectory and depends only on the initial and final positions.

We are interested in the work done *against* the field force Eq. (1.28). Integrating on the curve Γ from position \mathbf{r}_1 to position \mathbf{r}_2 , we have

$$W = - \int_{1,\Gamma}^2 \mathbf{F} \cdot d\mathbf{s} = -q \frac{q_0}{4\pi\epsilon_0} \int_{1,\Gamma}^2 \frac{1}{r^2} \mathbf{u}_r \cdot d\mathbf{s} = -q \frac{q_0}{4\pi\epsilon_0} \int_1^2 \frac{1}{r^2} dr.$$

Note carefully that in the last member, we have dr because the force is radial. The last integral is independent of the trajectory, which ultimately gives us

Fig. 1.7 Elements for the line integral of the field of a point charge



$$W = -q \frac{q_0}{4\pi\epsilon_0} \left(\frac{1}{r_1} - \frac{1}{r_2} \right). \quad (1.29)$$

We have shown that the electrostatic force generated by a point charge is conservative and we can thus define an electrostatic potential energy U . The potential difference between positions 2 and 1 is equal to the work to be done against the field force to move the charge q from point 1 to point 2, namely

$$W = U(\mathbf{r}_2) - U(\mathbf{r}_1). \quad (1.30)$$

As is always the case, the electrostatic potential energy is defined up to an additive constant. Namely, it is

$$U(\mathbf{r}) = q \frac{q_0}{4\pi\epsilon_0} \frac{1}{r} + \text{const.} \quad (1.31)$$

In this case, it is natural, but not necessary, to define the potential energy as being zero at infinite distance. With this choice, the additive constant is zero.

Consider now the force field of n point charges or of a continuous charge distribution, at rest in both cases. The electrostatic force on a charge q at a generic point P due to the distribution of charges is the resultant of the forces exerted in P by each charge of the distribution. All the forces being applied at the same point P , the work of the resultant is equal to the sum of the works of those forces. Each of them is due to a point charge; and we have already shown that its work is independent of the integration path. Such is also the sum of their works, which is the work of the resultant. In conclusion, the electrostatic force due to any distribution of charges at rest is conservative and we can define its potential energy.

The electrostatic force acting on the unitary charge is the electric field (electrostatic in this case, to be precise). The potential energy difference per unit charge is called the *electrostatic potential* difference. Explicitly, the electrostatic potential difference between two points \mathbf{r}_1 and \mathbf{r}_2 is the work we must do against the field force to move the unit charge from \mathbf{r}_1 to point \mathbf{r}_2 :

$$\phi(\mathbf{r}_2) - \phi(\mathbf{r}_1) = - \int_{\mathbf{r}_1}^{\mathbf{r}_2} \mathbf{E} \cdot d\mathbf{r}. \quad (1.32)$$

As the potential energy, the potential is defined up to an additive constant. The constant is fixed by defining the potential at an arbitrarily chosen point as zero. Let \mathbf{r}_0 be the position vector of this point. The potential in a generic position \mathbf{r} is the work to be done against the field force to move the unit charge from \mathbf{r}_0 to \mathbf{r} . To find it, we choose an arbitrary curve from \mathbf{r}_0 to \mathbf{r} , and along it, we calculate the integral as

$$\phi(\mathbf{r}) = - \int_{\mathbf{r}_0}^{\mathbf{r}} \mathbf{E} \cdot d\mathbf{r}. \quad (1.33)$$

In the particular case of a point charge q_0 , the potential difference is

$$\phi(\mathbf{r}_2) - \phi(\mathbf{r}_1) = \frac{q_0}{4\pi\epsilon_0 r_2} - \frac{q_0}{4\pi\epsilon_0 r_1}. \quad (1.34)$$

The potential of a point charge q_0 at rest in the origin of the axes is then

$$\phi(\mathbf{r}) = \frac{q_0}{4\pi\epsilon_0 r} + \text{const.} \quad (1.35)$$

The constant is zero if we choose the potential to be zero at infinite distance. With this choice, the potential at a point is the work to be done against the field to move the unit charge from infinite distance to that point.

The potential at point \mathbf{r}_0 due to the n charges q_1 in \mathbf{r}_1 , q_2 in \mathbf{r}_2 , q_3 in \mathbf{r}_3 , ..., q_n in \mathbf{r}_n is

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^n \frac{q_i}{r_{i0}} + \text{const.} \quad (1.36)$$

Similarly, the potential at point $\mathbf{r}_1 = (x_1, y_1, z_1)$ due to a continuous distribution of charge with density $\rho(x_2, y_2, z_2)$ is

$$\phi(\mathbf{r}_1) = \frac{1}{4\pi\epsilon_0} \int_V \frac{\rho(x_2, y_2, z_2)}{r_{21}} dV_2 + \text{const} \quad (1.37)$$

where $dV_2 = dx_2 dy_2 dz_2$.

In the latter cases as well, it is usually convenient, but not necessary, to define the potential at infinite as zero. The “const” in the above equations is then zero.

The electrostatic potential has the physical dimensions of an energy per unit charge. Its unit is the volt, named after Alessandro Volta (Italy, 1745–1827) and equal to one joule per coulomb, $1 \text{ V} = 1 \text{ J C}^{-1}$.

We have seen how we should calculate the potential differences once we know the electric field. We shall now look into how to express the field knowing the potential. This is an immediate consequence of the *gradient theorem* of vector calculus. The theorem states that, for every scalar field ϕ (satisfying the conditions specified by calculus that are always met in practice), the line integral of its gradient ($\nabla\phi$) on every curve from \mathbf{r}_1 to \mathbf{r}_2 is equal to the difference between the values of ϕ at the two extremes, namely

$$\phi(\mathbf{r}_2) - \phi(\mathbf{r}_1) = \int_{\mathbf{r}_1}^{\mathbf{r}_2} \nabla \phi \cdot d\mathbf{r} \quad (1.38)$$

The left-hand side of this equation and Eq. (1.32) are equal. In addition, the relations hold for arbitrary integration paths on their right-hand sides. Consequently, the integrands must be equal, namely

$$\mathbf{E} = -\nabla \phi \quad (1.39)$$

The electric field is the opposite of the gradient of the potential. Indeed, stating that a vector field is the gradient of a scalar function is equivalent to stating that the line integral of the vector depends only on the origin and the end of the line. As a matter of fact, there are two other ways of stating the same thing.

The line integral on every closed line (called circulation) of the vector is zero:

$$\oint \mathbf{E} \cdot d\mathbf{s} = 0. \quad (1.40)$$

The curl of the vector is identically zero (it should be known from the vector calculus that the curl of a gradient is identically zero):

$$\nabla \times \mathbf{E} = 0. \quad (1.41)$$

Equations (1.40) and (1.41) state the same thing, but the latter is local, stating that the curl of the field is zero at each point of the space, and the former is a global (or integral) expression.

Finally, we note that the electrostatic potential gives a description of the system equivalent to the electrostatic field. Being one quantity instead of three, the former is often easier to compute. However, what we measure are the forces, namely the field.

1.6 Generating Electrostatic Potential Differences

The experimental study of electric phenomena, including the majority of those classified as electrostatic, is made possible by the availability of physical sources of potential difference. Batteries are the most common sources of continuous voltage, as the potential difference is also called. A battery is a device able to store electric energy in chemical form.

The description of the chemical and physical processes at the basis of the operation of voltage sources is beyond the scope of this course, and neither shall we need it. We shall, however, give a few basic pieces of information here.

A battery generates a voltage between two metallic parts, called its *electrodes* or *poles*, and an electric current when those electrodes are connected by a conductor. The electrode of higher voltage is called *positive* while the other one is *negative*. In this chapter, we shall use the device simply to produce a voltage, while it shall be used to produce currents in the next.

Historically, the first batteries were developed by Alessandro Volta in the last decade of the XVIII century and made public in 1800. Volta found that two different metals immersed in an electrolytic solution develop a definite potential difference. The basic unit is a “cell”, which was originally made of two disks, one of copper and one of zinc, separated by a disk of felt soaked in a solution of sulfuric acid H_2SO_4 . The cell has a voltage of about 1 V, with the copper acting as the positive pole. Volta also found that overlapping n cells one on top of the other in a pile, namely disks of copper, electrolyte, zinc, copper, electrolyte, zinc, and so on, he could obtain a voltage n -times larger than he could with a single cell. This is the Volta pile. A picture of a reproduction of one of Volta’s original piles is housed in the History Museum of Pavia University and is shown in Fig. 1.8.

The pile opened the way to the study of electric currents and their effects, including magnetic fields. The technique was strongly developed in the subsequent years. Note that voltaic piles, and their similar successors, can only deliver a finite amount of charge, namely a temporary current. The current is produced by chemical reactions inside the pile. Rechargeable batteries were developed starting in the second half of the XIX century.

Today, a large number of different batteries have become available, from the heavy ones used in our cars to the light and very light ones in our laptops and cell phones. The technology is still one under important development.

Fig. 1.8 Alessandro Volta pile (reproduction by Gelside Guatterini 1999). Museo per la Storia dell’Università—Sistema Museale di Ateneo—Pavia University



1.7 Calculating Electrostatic Potentials

In this section, we shall calculate the potentials of two simple charge distributions. We have already calculated the corresponding fields in Sect. 1.4. These distributions extended to the infinite and are unrealistic, but they are useful in demonstrating a particular difficulty and how to deal with it.

Linear charge distribution.

Consider a linear distribution of uniform linear charge density λ . Let us calculate the potential difference between a generic point at a distance r' from the line and a reference point at a distance r'_0 . The field is given by Eq. (1.24). The potential difference is

$$\phi(r') - \phi(r'_0) = - \int_{r'_0}^{r'} \mathbf{E} \cdot d\mathbf{s} = - \frac{\lambda}{2\pi\epsilon_0} \int_{r'_0}^{r'} \frac{dr'}{r'} = - \frac{\lambda}{2\pi\epsilon_0} \ln r' + \frac{\lambda}{2\pi\epsilon_0} \ln r'_0. \quad (1.42)$$

The result looks strange at a first sight because the arguments of the logarithm functions should be, but are not, pure numbers. The consequence seems to be that, if we change the measurement unit of length, we obtain a different value, which is clearly absurd. The problem does not exist, however, because only potential differences are meaningful and because the difference between two logarithms is the logarithm of the ratio between their arguments, and that ratio is dimensionless. Indeed, we can write the above equation as

$$\phi(r') - \phi(r'_0) = - \frac{\lambda}{2\pi\epsilon_0} \ln \frac{r'}{r'_0}. \quad (1.43)$$

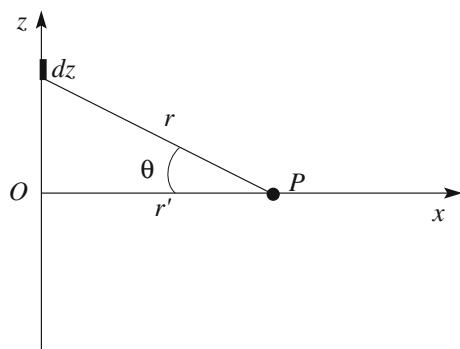
Looking back at Eq. (1.42), we observe that the additive constant of the potential, which is $\frac{\lambda}{2\pi\epsilon_0} \ln r'_0$, diverges when the reference point, namely r'_0 , goes to infinity. This means that the work to be done against the field force to move the unit charge from infinity to any point at finite distance from the line is infinitely large. This is because the force decreases very slowly when the distance increases, a consequence of the line being infinitely long. This never happens in practice. Indeed, the potential differences between points at finite distances are finite.

Let us now directly calculate the potential. As we can see in Fig. 1.9, the contribution to the potential of an infinitesimal segment dz is

$$d\phi = \frac{\lambda}{4\pi\epsilon_0} \frac{dz}{r}$$

where we have put the zero potential at infinity. This is completely safe for the segment dz . The problem arises when we sum up, namely we integrate on the line.

Fig. 1.9 Infinitesimal contribution to the electric potential of an element of a linear uniform charge distribution



To this purpose, let us express all terms as functions of θ , namely $r = r'/\cos \theta$, $z = r' \tan \theta$ and $dz = (r'/\cos^2 \theta)d\theta$, obtaining

$$d\phi = \frac{\lambda}{4\pi\epsilon_0} \frac{d\theta}{\cos \theta}.$$

We now integrate from $-\pi/2$ to $+\pi/2$, obtaining

$$\begin{aligned} \phi &= \frac{\lambda}{4\pi\epsilon_0} \int_{-\pi/2}^{+\pi/2} \frac{d\theta}{\cos \theta} = \frac{\lambda}{4\pi\epsilon_0} 2 \int_0^{+\pi/2} \frac{d\theta}{\cos \theta} \\ &= \frac{\lambda}{2\pi\epsilon_0} \left[\ln \tan(\theta/2 + \pi/4) \right]_0^{\pi/2} = \frac{\lambda}{2\pi\epsilon_0} [\ln \tan(\pi/2) - \ln \tan(\pi/4)] = \infty. \end{aligned}$$

The result cannot be used. However, in practice, one always deals with finite segments of line. The limits of the integral are never $+\pi/2$ to $-\pi/2$, there are no infinites, and the result makes sense.

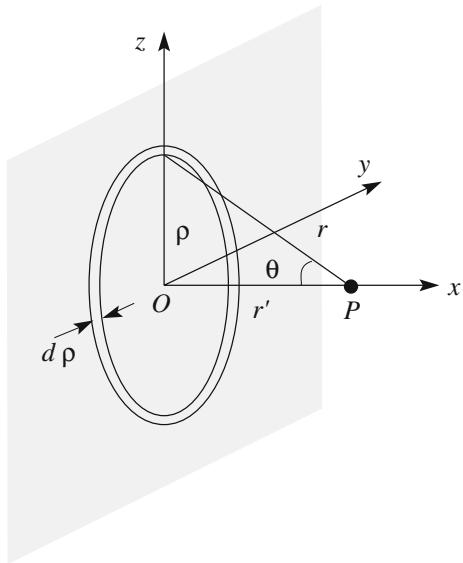
Planar charge distribution.

Let us now consider an infinite plane with uniform surface charge density σ . Let us try to calculate the potential at a generic point P at a distance r' from the plane, summing the infinitesimal contributions of surface elements in the form of annuluses, as shown in Fig. 1.10. We write the contribution of an element as

$$d\phi = \frac{\sigma}{4\pi\epsilon_0} 2\pi\rho d\rho \frac{1}{r},$$

which, once more, assumes zero potential at infinity. When we sum all the elementary contributions, we shall calculate the work needed to move the unit charge from infinity (the reference point) to a finite position. We already know that the result will be infinite, because the force is constant up to infinite distances. This is clearly a non-physical situation, but let us do the calculation anyway.

Fig. 1.10 Infinitesimal contribution to the electric potential of an element of a uniform planar charge distribution



We take into account that $r'^2 + \rho^2 = r^2$ with r' being fixed, and, consequently, that $\rho d\rho = r dr$. Then, we have

$$\phi(r') = \frac{\sigma}{2\epsilon_0} \int_{r'}^{\infty} \frac{\rho d\rho}{r} = \frac{\sigma}{2\epsilon_0} \int_{r'}^{\infty} dr = \infty - \frac{\sigma r'}{2\epsilon_0}.$$

This is the potential difference between a point at distance r' and one at infinite distance. As expected, it is infinite. Once more, this is because the charge distribution is unlimited, which never happens in practice. However, even in this case, the potential difference between two points at finite distances is finite. Let r'_1 and r'_2 be such distances. Their potential difference is

$$\phi(r'_2) - \phi(r'_1) = - \int_{r'_1}^{r'_2} \mathbf{E} \cdot d\mathbf{s} = - \frac{\sigma}{2\epsilon_0} \int_{r'_1}^{r'_2} dr' = - \frac{\sigma r'_2}{2\epsilon_0} + \frac{\sigma r'_1}{2\epsilon_0}$$

We see once more that the potential difference diverges when one of the two points moves to infinite distance. The above discussion was conducted to warn the reader that he/she might meet apparently absurd results, which, however, can be easily controlled. With this warning, we have no problem writing the potential of an infinite charged plane as

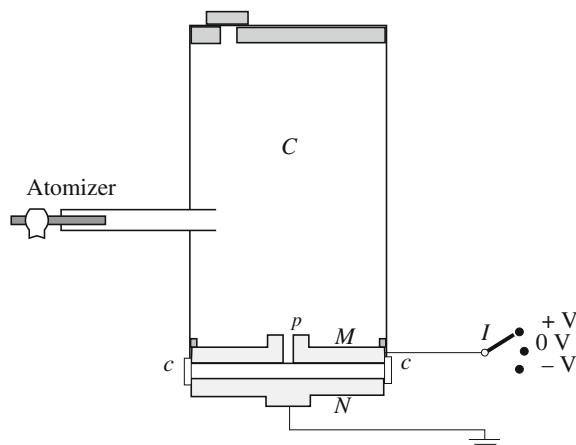
$$\phi(r') = - \frac{\sigma r'}{2\epsilon_0} + \text{const} \quad (1.44)$$

1.8 Measuring the Elementary Charge. Millikan's Experiment

Starting in 1909, Robert Millikan (USA; 1868–1953) developed and performed a series of elegant and precise experiments. He proved that electric charges are multiples of an elementary unit, called the elementary charge, which is the charge of the electron. He also accurately measured the electron charge. The electron itself had been discovered by Joseph John Thomson (UK, 1856–1940) in 1897. Millikan's method consists of the observation of small oil drops carrying a few elementary charges moving in an electric field. We shall take a historical point of view and describe the ingenious oil-drop experiment in some detail. We shall see, in particular, how the experimental demonstrations of the two statements above, while employing the same method, are largely independent of one another. We shall follow almost verbatim the description given by Millikan in the second edition of “The electron”, a small, beautiful book he published in 1924.

The device in its simplest form is shown schematically in Fig. 1.11. Using an atomizer, one blows oil droplets into the upper chamber C . The droplets are minuscule, with diameters of a few micrometers. They appear as a fog and slowly fall down. From time to time, one of them goes through the minuscule hole p made with the point of a pin in the middle of the plate M (22 cm in diameter). This plate and a second parallel one N form a capacitor. The two plates are taken at a distance of 16 mm by insulating spacers (three ebonite columns located along the circumference). Usually, the gas between the two plates is air at atmospheric pressure but different gases can be inserted if needed. Acting on the switch I , one can connect the plate M to a positive voltage source up to a few kilovolts, to a similar negative voltage, or to ground. Plate N is grounded. In this way, one can establish between the plates a vertical uniform electric field upward or downward directed of a few kV/cm or no field.

Fig. 1.11 Scheme of the Millikan oil-drop experiment



An intense light beam enters and exits through two small glass windows c through the volume in which the drop that has gone through the hole is moving. The drop is observed with a telescope through a third small window (not shown in the figure) at 90° with the light beam. The droplet scatters some light and appears to the experimenter as a bright star in a dark field.

The droplets are usually charged due to rubbing produced in the atomizer when the fog is blown. One can choose the direction and magnitude of the field to cause the droplet under observation to slowly rise toward M . Immediately before it reaches M , the switch I is moved to zero electric field and the droplet falls down under the action of its weight. Immediately before it reaches the plate N , the field is switched back on so that the droplet will rise again. In this way, one obtains a sequence of up and down motions of the same droplet. The motions are under the action of two “active” forces, the weight and the electrostatic force, and the air resistance. The latter is proportional to the velocity, which is the velocity for which the drag equilibrates the resultant active force.

Two horizontal reference lines separated by a known distance in the eyepiece of the telescope allow the observer to determine the velocities by measuring the times taken by the drop to move from one line to the other.

Table 1.1.1 features one of Millikan’s first measurement series. In this initial phase, he measured the times with a simple handheld stopwatch, with a probable error of a few tenths of a second. The real distance between the two reference lines was $d = 5.222$ mm.

The t_g row of the table shows the series of fall times. One sees that they are identical within the measurement errors. The t_F row reports the corresponding series of ascent times under the action of the electric field, whose magnitude was 3.157 kV/cm. One sees that, from the second to the third ascent, the time changes from 12.4 to 21.8 s. Millikan knew from the direction of the field that the drop was initially positive. The observed change thus meant that the droplet had captured a negative ion from the air during the second descent. Note that no ion can be captured when the field is on, because any ion present drifts immediately to one of the electrodes (depending on its sign). The following ascent time was 34.8 s, showing that a second negative ion had been captured. The next ascent time was 84.5 s, corresponding to the capture of a third negative ion. This charge was kept equal for two ascents. After them, the ascent time went back to 34.6 s, showing the capture of a positive ion. The newly captured positive ion had to have exactly the same charge, but with opposite sign, of the negative ion that had caused the opposite variation of the time (namely from 34.8 to 84.5 s).

The obvious objection to the above conclusions is that the drop might have lost an ion, instead of capturing one of the opposite sign. Millikan, however, showed that his droplets never lose ions with the following experiment. He measured the

Table 1.1 Oil drop ascent and descent times in a Millikan experiment

t_g (s)	13.6	13.8	13.4	13.4	13.6	13.6	13.7	13.5	13.5	13.8	13.7	13.8
t_F (s)	12.5	12.4	21.8	34.8	84.5	84.5	34.6	34.8	16.0	34.8	34.6	21.9

charge rate of change of drops while decreasing air pressure. He found the rate to be proportional to the pressure down to a few hectopascal. At this low pressure, a droplet could keep its charge invariable for several hours. The conclusion was that drops change charge by ion capture, not by ion emission.

From these simple measurements, we reach the extremely important conclusion that electric charge is always an integer multiple of a well-defined one, namely that electric charge is quantized. This conclusion is reached under the simple hypothesis that the droplet velocity is proportional to the active force (weight) when it is falling with no electric field and to the resultant of weight and electrostatic force when the field is on. The hypothesis implies the assumption of independence of the drag force of the droplet charge. Millikan took care to verify experimentally the correctness of these assumptions.

Let us go back to the data in Table 1.1 and indicate with w the droplet velocity in its descent. We can take the mean value of the measurements because they are consistent. Let us indicate with v , v' , etc., the ascent velocities. In the just made assumptions, w is proportional to mg (if m is the mass of the droplets) and each v is proportional to $qE - mg$, where q is the charge of the droplet in that ascent (do not confuse it with an ion charge). We have

$$\frac{w}{v} = \frac{mg}{qE - mg} \quad (1.45)$$

or, solving for q ,

$$q = \frac{mg}{wE} (w + v) \quad (1.46)$$

Consider when an ion capture changes the drop charge from a q to q' and, as a consequence, the speed from v to v' . The captured charge should be

$$\Delta q = q' - q = \frac{mg}{wE} (v' - v) \quad (1.47)$$

The factor mg/wE is constant for a given drop. Consequently, each captured charge is proportional to the change in the drop velocity ($v' - v$) in the electric field E . Note that we can reach this conclusion without knowing the field, the drop mass, etc. We can group the ascent times t_F in Table 1.1 in five sets, each with equal values within the errors. Each set corresponds to a different charge. We take the mean values of t_F , divide the known observation distance $d = 5.222$ mm by these mean values, and obtain the values in Table 1.2.

Table 1.2 reports, in the first column, the ascent velocities of the drop, and in the second column, the velocity change for an ion capture. We see that all the captured charges have the same value, within the errors, with three exceptions. In these three cases, the captured charge is exactly twice as large.

Table 1.2 Ascent velocities and descent velocity changes in a Millikan experiment

v (mm/s)	$v' - v$ (mm/s)
$5.222/12.45 = 0.4196$	
$5.222/21.85 = 0.2390$	$0.1806:2 = 0.0903$
$5.222/34.7 = 0.1505$	$0.0885:1 = 0.0885$
$5.222/85.0 = 0.06144$	$0.0891:1 = 0.0891$
$5.222/34.7 = 0.1505$	$0.0891:1 = 0.0891$
$5.222/16.0 = 0.3264$	$0.1759:2 = 0.0880$
$5.222/34.7 = 0.1505$	$0.1759:2 = 0.0880$
$5.222/21.85 = 0.2390$	$0.0891:1 = 0.0891$

Here, we have just given an example of the many similar measurements made by Millikan. He observed thousands of droplets of different sizes, some of them for several hours, in the presence of different gases at different pressures, with tens of thousands of ion captures. He increased the precision of the time measurements. The captured charges were always equal to the smallest charge or to a small integer multiple of it. He then concluded:

Here, then, is direct, unimpeachable proof that the electron is not a “statistical mean”, but that rather the electrical charges found on ions all have either exactly the same value or else small exact multiples of that value.

We note that Millikan’s method works because the electric force on a few elementary charges is comparable to the weight of a droplet. This happens because electric force is much larger than gravitational force. Let us look at the orders of magnitude. The oil droplets had diameters on the order of a micrometer, hence weights, as is immediately computed, of a few femtonewton. We see that the electric force on a few elementary charges in a field 10 kV/m has the same order of magnitude. Notice that, even if very small, a drop contains billions of molecules, while the number of ions on it is of a few units.

We can use the measurements of the descent speeds, i.e., when the field is not present, to obtain the initial charge of the droplet, produced by friction in the atomizer. Looking at Eqs. (1.46) and (1.47), we see that the charge of the droplet is proportional to $(w + v)$ by the same coefficient of proportionality as the captured charge is proportional to $(v' - v)$. Adding the mean value of the descent speed w to the speed of the first ascent v , we have $w + v = 0.8038$ mm/s. Dividing this value by 9, we have 0.0893 mm/s, which differs by less than 2 per mille from the values of the second column of Table 1.2. We conclude that the charge obtained by rubbing, in this case, is nine elementary charges. Millikan made many similar measurements with drops of different diameters and different materials, both conductors and insulators, reaching the same conclusion in all cases: electric charge is quantized. In his words:

Our experiment has then given us for the first time a means of comparing a frictional charge with the ionic charge, and the frictional charge has in this instance been found to contain exactly 9 electrons.

Let us now consider the measurement of the absolute value of the elementary charge. Let us start with the theoretical analysis of the droplet motion.

In the descent, the drop moves under the action of its weight and of the viscous drag. As learnt in the second volume of this course, the drag on a perfectly rigid and smooth sphere of the size and at the speeds we are considering is given by the Stokes law. If η is the viscosity of the medium, usually air, a the radius of the droplet and w its velocity, the drag force is

$$R = 6\pi aw\eta \quad (1.48)$$

This law, theoretically established by George Gabriel Stokes (Ireland and UK, 1819–1903), was experimentally and very accurately verified by Harold De Forest Arnold (USA, 1883–1933) (see Vol. 2, Sect. 1.14) for rigid spheres having radii between 10 μm and 1 mm. The Millikan droplets were smaller by about an order of magnitude, but let us assume the Stokes law to be valid for them.

During its fall, the droplet immediately reaches the regime speed, at which the resultant active force and the drag are equal and opposite. The active forces, to be precise, are the weight and the buoyancy (which we had neglected up to now as being irrelevant for the relative measurements). If ρ and ρ' are the densities of oil and of air, respectively, we have

$$\frac{4}{3}\pi a^3(\rho - \rho') = 6\pi a\eta w. \quad (1.49)$$

An additional force acts in the ascent, at constant velocity v . This is the electrostatic force, which is parallel and opposite to the weight. In this case, let us also write the equation of the equilibrium of the forces, taking into account that above, we have taken the speeds directed upward as positive. We have

$$qE - \frac{4}{3}\pi a^3(\rho - \rho') = 6\pi a\eta v. \quad (1.50)$$

In this equation, the electric field E is known, knowing the potential difference between the plates and their distance. We can also measure densities and viscosity. We obtain the radius a of the droplet by measuring the descent velocity and using Eq. (1.49). Equation (1.50) then gives us the drop charge q .

We must be very careful here, because we want to obtain the absolute value of a fundamental constant. The uncertainties of all the quantities of the game, densities, viscosity, electric field, etc., contribute (by error propagation) to the uncertainty of the final result. Viscosity, in particular, depends on temperature. In a second version of his experiment, Millikan submerged the apparatus in a bath of constant temperature within 0.02 °C. Under these conditions, he knew and could control viscosity within one per mille.

A further condition is the validity of the Stokes law. Millikan observed that the values of the elementary charge calculated under this assumption for drops falling at

the same velocity, which had equal radii, were equal within errors. Contrastingly, values were different for drops of different velocities, hence of different radii. The resulting differences were up to several percent. Millikan concluded that the Stokes law could not be completely valid. As a matter of fact, one of the assumptions made by Stokes is the homogeneity of the medium. The air can be considered to be such as long as we do not look at it at scales comparable to the distances between molecules. The Stokes law cannot hold when the radii of the droplets are comparable to the vacancies between the air molecules. One can think of the average size of these vacancies as being on the order of the mean free path l between collisions of the gas molecules. As we saw in Sect. 6.4 of Vol. 2 of this course, the mean free path is inversely proportional to pressure, being about 70 nm at STP. The relative values of the ratio between Millikan's droplet radii and the mean free path were quite small, namely $l/a = 0.1 - 0.01$. Millikan performed a series of very accurate experiments dedicated to the measurement of the corrections to be applied to the Stokes law under such conditions. In 1916, he published the first measurement of the elementary charge:

$$q_e = 1.592 \pm 0.0016 \times 10^{-19} \text{ C.} \quad (1.51)$$

The accuracy was already of 10^{-3} . Today, this fundamental constant is known with a 6 ppb (ppb = part per billion) accuracy:

$$q_e = 1.6021766208 \pm 0.0000000098 \times 10^{-19} \text{ C.} \quad (1.52)$$

QUESTION Q 1.3. Find the velocity given by Eq. (1.50) of a droplet of $20 \mu\text{m}$ radius in a field of 10 kV/m . \square

1.9 Invariance of the Electric Charge

We have already seen two fundamental properties of the electric charge; conservation and quantization. The third property is invariance, namely the electric charge of an object being independent of its velocity, whatever it may be. Considering that an object in motion appears to be at rest to an observer moving at its velocity, the just made statement is equivalent to saying that the electric charge is *invariant* in the transformation between a (inertial) reference S and another (inertial) one S' moving at any velocity \mathbf{v} relative to S . Clearly, no confusion can be made between charge invariance and charge conservation. Electric charge and speed of light are the two fundamental invariants of physics.

Let us look at a few of the simplest experimental proofs of electric charge invariance. We shall take advantage of the perfect equality, in absolute value, of the proton and electron charges, which we already mentioned in Sect. 1.2.

Let us consider a hydrogen molecule, which is composed of two hydrogen atoms. Each of them is made of a proton and an electron. To establish if the

molecule is globally neutral, we must measure its charge. We shall never be able to state that it is exactly zero, due to the experimental errors, but we shall rather find it to be smaller than a certain value determined by the experimental sensitivity.

Let δq be the absolute value of the difference between the absolute values of the electron and proton charge. A direct measurement of δq is done by producing a beam of H₂ molecules, projecting it through an intense electric field and looking for any deflection of the beam. In an experiment of this type conducted in 1963, J.C. Zorn (USA, born 1931), G.E. Chamberlain and V.W. Huges (USA, 1921–2003)¹ were able to set the limit $\delta q < 10^{-15} q_e$, which is very small indeed. Even more sensitive are the experiments using beams of alkali atoms, both because they are easier to detect (see Chap. 5 of the second Volume) and because such atoms contain many more protons and electrons. The same authors working with a Cs beam obtained the limit of $\delta q < 10^{-18} q_e$.

An even more sensitive method, but somewhat indirect, being on a macroscopic system, consists of letting a gas escape from an electrically-insulated metal container. If each molecule had even an extremely small charge, the container would gradually build up a charge as the gas exits. This charge is measured with an electrometer. The method is, in principle, extremely sensitive, because it deals with some 10^{22} molecules. Clearly, the presence of even a very small number of ions must be avoided. This can be done by properly de-ionizing the gas. In an experiment of this type in 1960, J.G. King (USA, 1925–2014)² established the neutrality of both hydrogen molecules and helium atoms, obtaining limits of a few $10^{-20} q_e$.

Let us examine the consequences of both hydrogen molecules and helium atoms being neutral. Both of them contain two protons and two electrons (we ignore the two neutrons of helium, which are neutral). However, in the molecule, protons and electrons have “atomic” kinetic energies on the order of the electronvolt. In helium, the electrons have similar energies, but the protons are inside the nucleus, where the kinetic energies are on the order of the MeV, namely a million times larger. In the two systems, the electrons’ velocities are similar, while those of the protons are very different. The neutrality of both systems proves that the proton charge does not vary when its velocity changes by three orders of magnitude.

1.10 The Solid Angle

This section deals with a few geometric concepts that we shall need in the subsequent sections. We shall define the solid angle and its infinitesimal element. We shall see the relation between the latter and the volume element in polar co-ordinates.

Let us recall that, in planar geometry, an angle is the portion of a plane included between two half-lines, called sides, starting from a common point, say O , called

¹C. Zorn, G.E. Chamberlain and V.W. Huges; *Phys. Rev.* **129**, 2566 (1963).

²J.G. King *Phys. Rev. Lett.* **5**, 562 (1960).

the vertex. Let us take a circle with center in O and any radius R and let L be the length of the arc cut by the two sides of the angle on this circumference. The ratio L/R is independent of R and depends only on the amplitude of the angle. For this reason, this ratio is taken as the measure of the angle, namely $\phi = L/R$. The unit is the *radian* (rad).

Let $d\phi$ be an infinitesimal angle with vertex O , dL the arc of the circle of radius \mathbf{R} cut by the sides of this angle and \mathbf{u}_R the unit vector of \mathbf{R} . Consider also the arc dl of an arbitrary curve cut by the sides of $d\phi$ and let \mathbf{n} be the unit vector normal to dl , as shown in Fig. 1.12. Let θ be the angle between dl and dL , which is also the angle between \mathbf{n} and \mathbf{u}_R . Clearly, it is $dL = dl \cos\theta$ and we can write the following equivalent expressions of $d\phi$

$$d\phi = \frac{dL}{R} = \frac{dl \cos\theta}{R} = \frac{\mathbf{n} \cdot \mathbf{u}_R dl}{R}. \quad (1.53)$$

Let us now consider a point O in three dimensions. Take an arbitrary, but not through O , closed regular curve and all the half-lines from O touching the points of this curve. The space region defined by these lines is a *solid angle*. In other words, a solid angle is a cone of arbitrary section.

Consider a spherical surface centered in O of arbitrary radius R . The area S of the spherical cap cut by the solid angle on this surface is proportional to R^2 , and we can take as the measure of the solid angle the ratio, which is independent of R , between the area of any spherical cap with center in O and the square of its radius, namely

$$\Omega = \frac{S}{R^2}. \quad (1.54)$$

The unit of the solid angle is the *steradian* (sr). The complete solid angle measures 4π sr.

Figure 1.13 shows an elementary solid angle. It cuts two surfaces: dS , which is part of the sphere centered in O of radius R , and $d\Sigma$, which is arbitrary. Let \mathbf{n} be the unit vector normal to dS , \mathbf{u}_R the unit vector of the radius \mathbf{R} , which is normal to dS ,

Fig. 1.12 Planar elementary angle

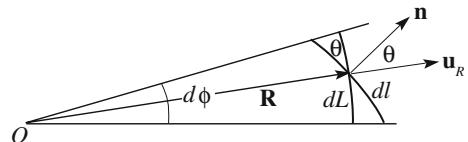
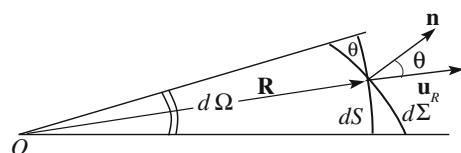


Fig. 1.13 Solid elementary angle



and θ the angle between them. dS being normal to the radius, the elementary solid angle is $d\Omega = dS/R^2$. If we need it in terms of $d\Sigma$, we must take the projection of this surface as being normal to the radius. Equivalent expressions of the elementary solid angle are

$$d\Omega = \frac{dS}{R^2} = \frac{d\Sigma \cos \theta}{R^2} = \frac{\mathbf{n} \cdot \mathbf{u}_R d\Sigma}{R^2}. \quad (1.55)$$

The solid angle under which a finite surface Σ is seen from O is obtained by integration on that surface, namely

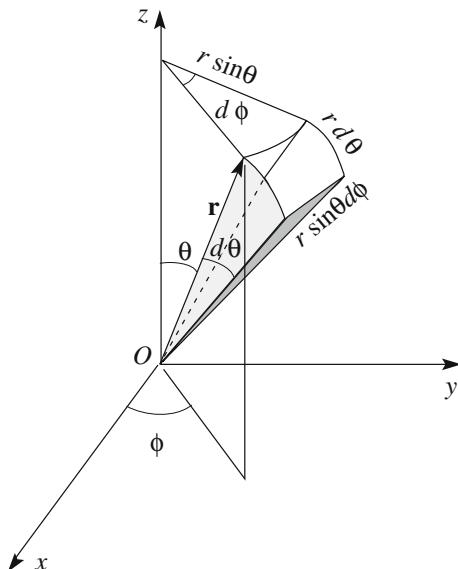
$$\Omega = \int_{\Sigma} d\Omega = \int_{\Sigma} \frac{\mathbf{n} \cdot \mathbf{u}_R d\Sigma}{R^2}. \quad (1.56)$$

The solid angle element Eq. (1.55) is immediately linked to the volume element in polar co-ordinates. A system of polar co-ordinates (r, θ, ϕ) is shown in Fig. 1.14.

It is not infrequent to encounter problems that have spherical symmetry. In these cases, it is usually convenient to use polar co-ordinates, rather than Cartesian co-ordinates. When a volume integration is needed, what is the expression of the elementary volume element dV ? In Cartesian co-ordinates, this is simply the product of the differentials of the co-ordinates, namely $dV = dx dy dz$. In polar co-ordinates, it is not so simple.

Let us determine the area drawn by the position vector \mathbf{r} when the polar angle varies from θ to $\theta + d\theta$ and the azimuth varies from ϕ to $\phi + d\phi$. The tip of

Fig. 1.14 Volume element in polar co-ordinates



\mathbf{r} moves along the arc $r d\theta$ in the polar angle variation and along the arc $r \sin\theta d\phi$ in the azimuth variation. The two displacements being perpendicular, the area is $dS = r^2 \sin\theta d\theta d\phi$. This area is normal to the position vector and is consequently equal to the solid angle $d\Omega$ under which it is seen from O multiplied by the distance square r^2 . In conclusion, the solid angle element is

$$d\Omega = \sin\theta d\theta d\phi. \quad (1.57)$$

To obtain the volume element, we must now vary the third co-ordinate, namely the position vector, from r to $r + dr$. We obtain a volume of base dS and height dr . Hence, the volume element is

$$dV = r^2 dr \sin\theta d\theta d\phi = r^2 dr d\Omega. \quad (1.58)$$

QUESTION Q 1.4. Evaluate the solid angle under which you see the fingertip of your thumb with your arm outstretched. \square

1.11 The Flux of \mathbf{E} and the Gauss Theorem

Let $\mathbf{E}(\mathbf{r})$ be an arbitrary (regular) vector field, $d\Sigma$ an oriented surface element and \mathbf{n} the unit vector of its positive normal. We define as the *flux* of the vector \mathbf{E} through $d\Sigma$ the quantity

$$d\Phi = \mathbf{E} \cdot \mathbf{n} d\Sigma. \quad (1.59)$$

The reason for the name is clear. If \mathbf{E} is the velocity of an incompressible fluid, the above-defined flux is the volume of fluid flowing through the surface $d\Sigma$ in a second.

The flux through a finite surface Σ is obtained by integration:

$$\Phi = \int_{\Sigma} d\Phi = \int_{\Sigma} \mathbf{E} \cdot \mathbf{n} d\Sigma. \quad (1.60)$$

The flux of \mathbf{E} through Σ is the integral on Σ of the component of \mathbf{E} normal to Σ .

The just given definition holds for any vector field. We now consider the electric field \mathbf{E} . We start from the field of a point charge q . Let Σ be a closed surface enclosing q oriented with the positive unit vector outward, as in Fig. 1.15a. Let us express the flux of the electric field of q through Σ . To be precise, this is the *outgoing* flux, given the choice we made for the orientation of Σ . Let \mathbf{r} be the position vector drawn from q and \mathbf{u}_r its unit vector. The flux through the surface element $d\Sigma$ is

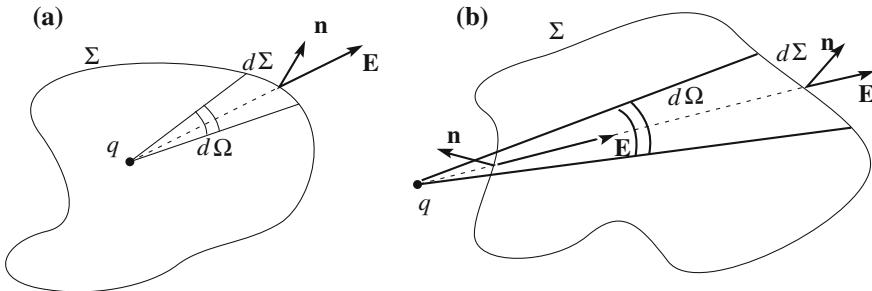


Fig. 1.15 Flux of electric field of a point-charge **a** inside, **b** outside the surface

$$d\Phi = \mathbf{E} \cdot \mathbf{n} d\Sigma = \frac{q}{4\pi\epsilon_0 r^2} \mathbf{u}_r \cdot \mathbf{n} d\Sigma = \frac{q}{4\pi\epsilon_0} d\Omega. \quad (1.61)$$

where we have used Eq. (1.55). We see that the flux of the field depends only on the solid angle under which the charge “sees” the surface element. This is due to two properties of the electric field of a point charge at rest: it is radial and its magnitude is inversely proportional to the distance squared. As such, the field decreases exactly as the subtended area increases and the two variations cancel one another out. Even the smallest difference from -2 in the exponent of r would destroy the cancellation.

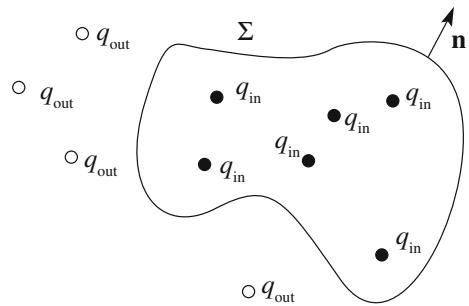
The outgoing flux through the whole surface is obtained by integration, namely it is

$$\Phi = \int_{\Sigma} \mathbf{E} \cdot \mathbf{n} d\Sigma = \frac{q}{4\pi\epsilon_0} \int_{\text{all}} d\Omega = \frac{q}{\epsilon_0}. \quad (1.62)$$

The flux of the electric field of a point charge at rest through any closed surface enclosing the charge is equal to that charge divided by ϵ_0 , independent of the position of the charge, provided it is inside. We shall soon see that this result holds for any number of charges. Before doing that, let us express the flux through a closed surface of the field of a point charge outside the surface, as in Fig. 1.15b. As we see in the figure, every solid angle element $d\Omega$ cuts *two* surface elements (or, more generally, an even number of elements). The fluxes through them are equal in magnitude, for what we have just seen, and opposite in sign, because the outside normal to the surface has opposite directions relative to the field. As a result, the flux elements cancel in pairs and the total result is zero.

Consider now the electric field produced by an arbitrary distribution of charges and a closed surface Σ , as in Fig. 1.16. The flux of the electric field outgoing from Σ is simply the sum of the outgoing fluxes of each of the charges, because the superposition principle holds for the field and because integration is a linear operation. Each of the internal charges q_i contributes to the flux with q_i/ϵ_0 . All the external charges give zero contributions. In conclusion, calling Q_{in} the sum of the internal charges, the flux of the electric field of the distribution is

Fig. 1.16 A charge distribution and a closed surface



$$\Phi = \frac{Q_{in}}{\epsilon_0}. \quad (1.63)$$

This is the Gauss law, after the great German mathematician Carl Friederich Gauss. The law states that *the flux of the electric field outgoing from a closed surface is equal to the charge inside that surface divided by ϵ_0* .

Suppose now that we can consider the charge to be continuously distributed with density ρ , which is, in general, a function of the position. In this case, the charge inside the surface is the volume integral of ρ over the volume V inside the surface, and we have

$$\Phi = \int_{\Sigma} \mathbf{E} \cdot \mathbf{n} d\Sigma = \frac{1}{\epsilon_0} \int_V \rho dV. \quad (1.64)$$

This is the Gauss law in integral form. We can also express it in an equivalent differential form. We do that using a theorem of vector calculus called the divergence theorem, which is also credited to Gauss. The theorem states that the volume integral of the divergence of a vector field on volume V is equal to the flux of that vector outgoing from the surface Σ surrounding that volume, namely that

$$\int_V \nabla \cdot \mathbf{E} dV = \int_{\Sigma} \mathbf{E} \cdot \mathbf{n} d\Sigma. \quad (1.65)$$

Equation (1.64) becomes

$$\int_V \nabla \cdot \mathbf{E} dV = \frac{1}{\epsilon_0} \int_V \rho dV.$$

Being that this is true for every volume V , the equality must hold for the integrands and, ultimately, we have

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}. \quad (1.66)$$

This fundamental equation is equivalent, as we said, to Eq. (1.64). However, it is a *local relationship*. It establishes the equality between its sides at each space point. To know the divergence of the field at a point, we need only to know the charge density at that point.

The properties of the electric field we have established under static conditions in this section also hold under dynamic conditions, as we shall see in subsequent chapters.

1.12 Graphic Representation of the Electric Field

The electric field is mathematically a vector field. A vector field is a vector that is a function of the space point and, in general, of time. If it does not depend on time, as in the case we are considering now, the field is said to be static. The three components of the electric field are continuous and differentiable, singular points apart, as we shall see immediately. Equations (1.41) and (1.66) give the curl and the divergence of the field at every space point. Vector calculus teaches that the knowledge of curl, divergence and boundary conditions completely define a vector field. The divergence of the electrostatic field is continuous where the charge density is continuous. Such is also the case with the field, its curl always being zero, hence continuous. If point charges are present, in their position, the charge density, and hence the divergence of the electric field, are infinite. These are singular points of the field.

We shall now describe a graphic representation of the vector fields that is useful for visualizing the field, within certain limits, using lines of force and equipotential surfaces. We considered this exact representation in Sect. 4.8 of the first volume for the gravitational field, but we shall repeat the description here. The gravitational field is extremely similar to the electrostatic one, with the exception that its “charges”, the gravitational masses, have only one polarity, namely the gravitational field has sources but no sinks.

A *line of force* is an oriented curve having at every space point the positive direction of the field at that point. The lines of force are infinite in number, but there is only one of them at each point, singular points apart. In other words, the lines of force never cross each other; otherwise the field would have two directions at the crossing point. Beyond direction, the line of force can also represent the field intensity if we draw a number of lines such that their density at a point (namely the number of lines crossing the unitary surface normal to them) is proportional to the field intensity at that point.

The lines of force of the electrostatic field are continuous everywhere the field is continuous. They exit from the positive charges, which are called *sources* of the field, and enter the negative charges, called the *sinks*. The lines cannot start nor end at a point where there is no charge. This is a consequence of the inverse square dependence from distance of the field of a point charge. Figure 1.17 shows the lines of force of the field of a point charge Q . In this example, the charge is positive, and the lines exit from it because the field points out. They are straight. This guarantees

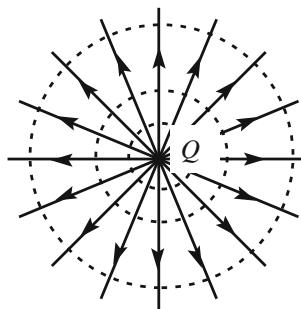


Fig. 1.17 The lines of force of the electrostatic field of a point positive charge

that numbers of lines crossing all the spheres of any radius centered on the charge are equal. Consequently, the lines cannot end or start outside the charge. We shall take the convention that the number of lines crossing a surface is equal to the flux of the electric field through that surface. Then, the number of lines coming out from Q if positive, or going into it if negative, should be Q/ϵ_0 .

Figure 1.18a shows graphically the field of two equal and opposite point charges, Fig. 1.18b of two equal (positive) charges. In the first case, all the lines leaving from the positive charge end in the negative one, with the exception of the line on the axis, which goes to infinity, and comes back on the other side. Such a pair of equal and opposite charges is called a dipole. In Fig. 1.18b, there is no negative charge and all the lines go to infinity. If two charges have opposite signs but are different in magnitude, for example, $+3q$ and $-q$, a fraction of the lines out of the positive charge ($1/3$ in this case) ends in the negative charge, while the rest goes to infinity.

It is also useful to include in the same graphic representation the potential, by the *equipotential surfaces*. An equipotential surface is the locus of points at which the potential have the same value. The surfaces, one for every potential value, are infinite in number. One chooses a value for the difference between two consecutive

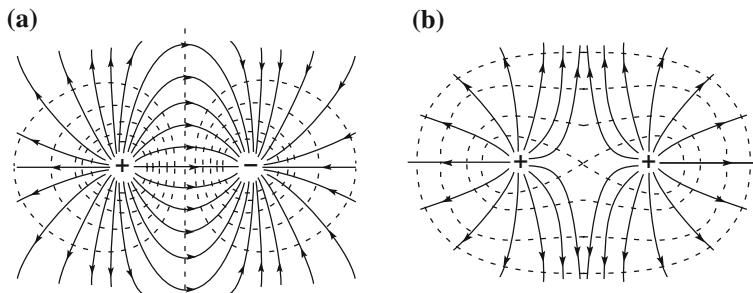


Fig. 1.18 Lines of field (*continuous*) and equipotential surfaces (*dotted*) for **a** two equal and opposite charges (a dipole), **b** to equal positive charges

surfaces, for example, one volt, and draws the surfaces at such a step. The equipotential surfaces (the dotted lines in Fig. 1.18) are perpendicular to the field, and hence to the field lines, at every point. Indeed, the field is the opposite of the gradient of the potential and the latter has zero components on the equipotential. The field intensity is larger where the equipotentials are closer to one another.

Notice in Fig. 1.18b the 8-shaped section of an equipotential. The crossing point, which is halfway between the charges, is a *saddle point* for the potential. Indeed, the equipotential surface there is shaped like a saddle. Moving in the direction of the charges, the potential increases on both sides, while moving perpendicularly to the line joining the charges, the potential decreases on both sides. The curve obtained by cutting the equipotential with a plane through the charges has a minimum at the saddle point, while the curve cut by a perpendicular plane has a maximum there.

The lines of force of the electrostatic field cannot be closed. This is a consequence of the curl of the field being identically zero. Indeed, if there were a closed line, the circulation integral about that line of the electric field would not be zero.

In the following section, we shall encounter fields lacking the latter property. Such are the magnetic field and the electric field itself in dynamic conditions.

1.13 Applications of the Gauss Law

The fundamental properties of the electrostatic field, as we have seen, are that its circulation around any closed line is zero and its flux outgoing from any closed surface is equal to the charge inside divided by ϵ_0 . Alternatively, we can say that the curl of the electric field is always zero and its divergence is the charge density divided by ϵ_0 . The solution to the problem of finding the field, for a given charge density, generally requires the use of both properties. However, particularly simple and symmetric charge distributions exist for which the problem can be solved using the Gauss law alone, with a bit of intuition. We shall now see some examples of that. In Chap. 2, we shall see the most important applications of the Gauss law, which are the properties of the conductors.

Charged spherical shell.

Consider a spherical surface with center at O and radius R . Let the charge q be distributed on the surface with uniform density. This is an idealized situation, because all the real charge distributions have a thickness, which might be small, but is never zero. Let us calculate the electric field of our surface.

Let us start from a point outside the sphere, at a distance r ($r > R$) from its center. Having the source spherical symmetry, the solution must have spherical symmetry as well. Consequently, the magnitude of the field shall depend only on the distance from the center, and its direction shall be the direction of the radius. Let $E(r)$ be the field magnitude.

Let Σ_1 be a spherical surface centered at O through the considered point, as in Fig. 1.19, and let us apply to that the Gauss theorem. The flux calculation is immediate, noticing that the field is perpendicular to the surface at all its points and its magnitude is equal at all of them. We have

$$\Phi = \int_{\Sigma_1} \mathbf{E} \cdot \mathbf{n} = E(r)4\pi r^2 = \frac{q}{\epsilon_0}, \quad (1.67)$$

from which we find the field. Let us include in its expression the direction that we know to be radial:

$$\mathbf{E}(\mathbf{r}) = \frac{q}{4\pi\epsilon_0 r^2} \mathbf{u}_r \quad \text{for } r > R. \quad (1.68)$$

We see that, outside the sphere, the field is equal to that which would occur if all the charge were concentrated in the center.

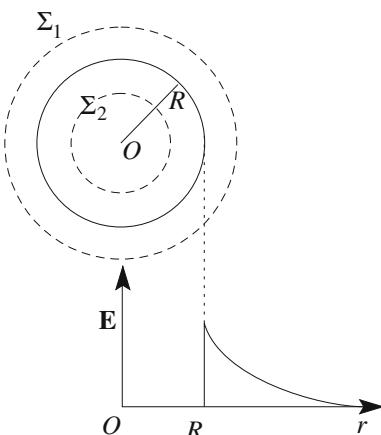
Consider now a point inside the sphere, namely at the distance from center $r < R$. We now once again apply the Gauss law to a spherical surface through the point that is now inside the sphere (Σ_2 in Fig. 1.19). The outgoing flux is zero, because the charge inside Σ_2 is zero. In addition, the symmetry of the problem requires the field to be equal at all the points along the surface. Hence, the field inside is zero.

$$\mathbf{E}(\mathbf{r}) = 0 \quad \text{for } r < R. \quad (1.69)$$

The reader may remember from the first volume that Eqs. (1.68) and (1.69) also hold for the gravitational field. Indeed, both fields decrease as the inverse square distance.

We now notice that the surface charge density is $\sigma = q/(4\pi R^2)$ and that Eq. (1.68) gives us on the surface of the sphere, namely for $r = R$,

Fig. 1.19 Electric field of a charged spherical shell



$$\mathbf{E}(\mathbf{r} = \mathbf{R}) = \frac{\sigma}{\epsilon_0} \mathbf{u}_r. \quad (1.70)$$

Crossing the surface, the electric field has a discontinuity equal to σ/ϵ_0 . We shall immediately see other examples of the same behavior.

Charged sphere.

Consider a solid sphere of radius R uniformly charged with total charge q .

We reason as in the previous case. The geometry is shown in Fig. 1.20. Outside the sphere ($r > R$), Eq. (1.68) still holds, because the symmetry of the problem is the same and the charge inside the surface is also the same. Inside the sphere, for $r < R$, we apply the Gauss law to the sphere Σ_2 in Fig. 1.20. The uniform charge density is $\rho = q/[(4/3)\pi R^3]$ and we have

$$\Phi = \int_{\Sigma_2} \mathbf{E} \cdot \mathbf{n} = E(r)4\pi r^2 = \frac{1}{\epsilon_0} \int_V \rho dV = \frac{q}{\epsilon_0} \frac{r^3}{R^3}$$

and then

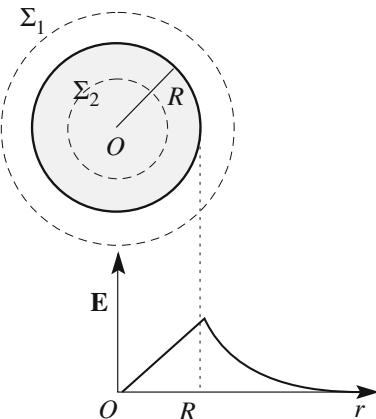
$$\mathbf{E}(\mathbf{r}) = \frac{q}{4\pi\epsilon_0 R^3} r \mathbf{u}_r \quad \text{for } r < R. \quad (1.71)$$

Inside the sphere, the field increases linearly with the distance from center. On the surface, Eqs. (1.71) and (1.68) give the same value. As opposed to the charged shell, the field is continuous.

Two spherical charged surfaces

A third case in which we can use the same arguments is the case of two charged surfaces, both centered at O . Let R_1 and R_2 ($R_1 < R_2$) be their radii. Let the two charges be equal and opposite, $+q$ on the inner sphere and $-q$ on the outer one.

Fig. 1.20 Electric field of an uniformly charged sphere



The argument now leads to the conclusions that the field is zero both outside the outer sphere ($r > R_2$) and inside the inner one ($r < R_1$), because, in both cases, the total charge contained on the spherical surface is zero. Between the charged spheres ($R_1 < r < R_2$), the field depends only on the charge of the inner one and is equal to that which would occur if the entire charge were concentrated in the center. In conclusion,

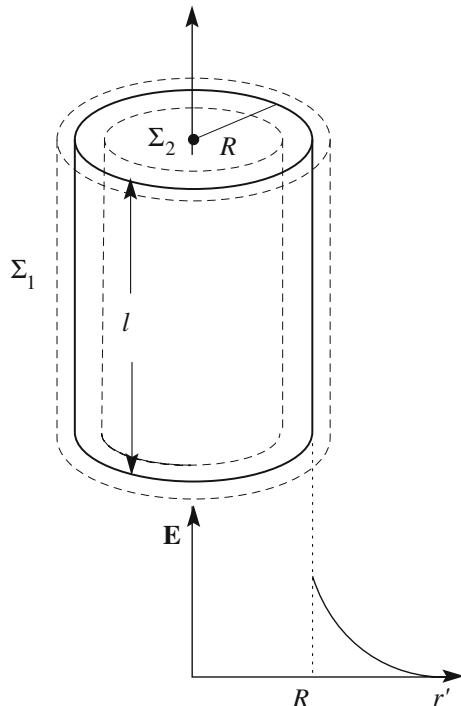
$$\mathbf{E}(\mathbf{r}) = \frac{q}{4\pi\epsilon_0 r} \mathbf{u}_r \quad \text{for } R_1 < r < R_2; \quad \mathbf{E}(\mathbf{r}) = 0 \quad \text{for } r < R_1 \text{ and } r > R_2. \quad (1.72)$$

Charged cylindrical surface

Consider an infinite cylindrical surface having z as axis and radius R uniformly charged with surface charge density σ . Let r' be the distance from the axis. Figure 1.21 shows the geometry.

We start with a point P outside the surface, namely at $r' > R$. We apply the Gauss law to the cylindrical surface Σ_1 coaxial to the charged cylinder through P (hence radius r') and height l , as shown in Fig. 1.21. For the symmetry of the problem, the field should have a magnitude dependent only on the distance from the axis (r') and be directed radially and perpendicularly to the axis. Hence, we have

Fig. 1.21 Electric field of a cylindrical shell



$$\Phi = \int_{\Sigma_1} \mathbf{E} \cdot \mathbf{n} d\Sigma = 2\pi r' l E(r') = 2\pi R l \frac{\sigma}{\epsilon_0}$$

and

$$\mathbf{E}(r') = \frac{\sigma R}{\epsilon_0 r'} \mathbf{u}_{r'} \quad \text{for } r' > R. \quad (1.73)$$

We can also express the field as a function of the linear charge density, namely the charge per unit length of the cylinder, and call it $\lambda = 2\pi R \sigma$. We have

$$\mathbf{E}(r') = \frac{\lambda}{2\pi\epsilon_0 r'} \mathbf{u}_{r'} \quad \text{for } r' > R, \quad (1.74)$$

which we recognize to be equal to the field of a linear charge density on the axis of the cylinder in Eq. (1.24).

Consider now a point inside the surface, namely at $r' < R$. We apply the Gauss law to the cylindrical surface Σ_2 coaxial to the charged cylinder through it (hence radius r') and height l , as shown in Fig. 1.21. By the now familiar symmetry argument, and because the charge inside the surface is zero, we have

$$\mathbf{E}(r') = 0 \quad \text{for } r' < R. \quad (1.75)$$

The field (which is perpendicular to the surface) has the discontinuity σ/ϵ_0 at the charged surface.

Solid charged cylinder.

Consider now a cylinder with radius R uniformly charged with charge density ρ . The corresponding charge per unit axis length λ is given by $\lambda = \pi R^2 \rho$. The field outside the cylinder is given by the same arguments as for the cylindrical surface, and the result is the same. If we express it in terms of the charge per unit length, we again have

$$\mathbf{E}(r') = \frac{\lambda}{2\pi\epsilon_0 r'} \mathbf{u}_{r'} \quad \text{for } r' > R. \quad (1.76)$$

To have the field inside the cylinder, at the distance from the axis $r' < R$, we consider the usual cylindrical surface Σ_2 with radius r' and height l . The Gauss law says that

$$\Phi = \int_{\Sigma_2} \mathbf{E} \cdot \mathbf{n} d\Sigma = 2\pi r' l E(r') = \frac{\rho \pi r'^2 l}{\epsilon_0} = \frac{\lambda \pi l}{\pi R^2 \epsilon_0} r'^2,$$

which gives us

$$\mathbf{E}(r') = \frac{\lambda r'}{2\pi\epsilon_0 R^2} \mathbf{u}_{r'} \quad \text{for } r' < R. \quad (1.77)$$

We immediately see that, on the surface of the cylinder, namely for $r' = R$, Eqs. (1.75) and (1.76) give the same value. The field is continuous.

Two cylindrical charged surfaces.

Consider two coaxial cylindrical charged surfaces. Let R_1 and R_2 ($R_1 < R_2$) be their radii. Let the two surface charge densities be equal and opposite, $+\sigma$ on the inner surface and $-\sigma$ on the outer one.

Once again applying the Gauss law, you find that the field is zero both inside the inner cylinder (for $r' < R_1$) and outside the outer one (for $r' > R_2$). Indeed, in both cases, the charge on the Gauss surface is zero. Between the two surfaces, the field is due to the inner one only and is given by Eq. (1.74), namely

$$\mathbf{E}(r') = \frac{\lambda}{2\pi\epsilon_0 r'} \mathbf{u}_{r'} \quad \text{for } R_1 < r' < R_2; \quad \mathbf{E}(r') = 0 \quad \text{for } r' < R_1 \text{ and for } r' > R_2. \quad (1.78)$$

Charged plane.

Consider an indefinite planar distribution of uniform surface charge density σ . Let P be the point at which we want the field and x a coordinate axis perpendicular to the plane through P . We apply the Gauss law to a cylindrical surface Σ of faces parallel to the charged surface, one through P , the other symmetrically. The height of the cylinder is $2x$ (Fig. 1.22).

For the symmetry of the problem, the field must be directed perpendicularly to the charged plane, outward if the charge is positive, inward if negative. In both cases, the directions on the two sides are opposite. The magnitude of the field can depend only on the distance from the plane x , if any. Applying the Gauss law, we have

Fig. 1.22 Electric field of a uniformly charged plane

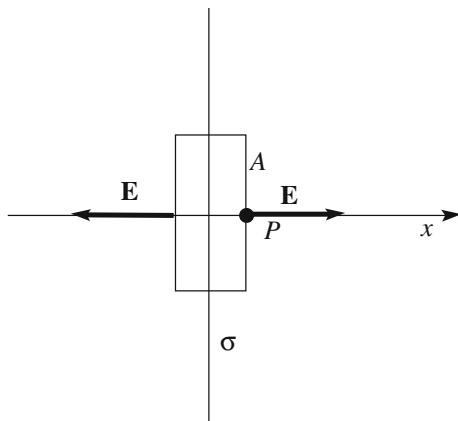
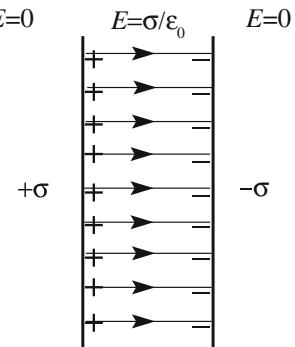


Fig. 1.23 The field of two parallel oppositely charged planes



$$\Phi = \int_{\Sigma} \mathbf{E} \cdot \mathbf{n} d\Sigma = 2E(x)A = \frac{\sigma A}{\epsilon_0},$$

which gives us, taking the sign into account,

$$E_x(x) = \frac{\sigma}{2\epsilon_0} \quad \text{for } x > 0; \quad E_x(x) = -\frac{\sigma}{2\epsilon_0} \quad \text{for } x < 0, \quad (1.79)$$

which is the result, independent of x , that we already obtained in Sect. 1.4. We note that the discontinuity of the perpendicular component of the field (the only existing one) in crossing the charged plane is, once more, σ/ϵ_0 .

Double plane

Consider two parallel planar distributions with uniform equal and opposite surface charge densities $\pm\sigma$ at a distance that, as we shall see, we do not need to specify, as in Fig. 1.23.

Applying the superposition principle to the just found result, we find that the field is zero outside the region between the plane (where the contributions of the two planes are equal and opposite). Between the planes, the field is twice the field of a single plane in between, because the contributions are equal in magnitude and sign, namely

$$E = \frac{\sigma}{\epsilon_0} \quad \text{directed from } +\sigma \text{ to } -\sigma, \quad (1.80)$$

Once more, the discontinuities of the perpendicular field components at the charged surfaces are σ/ϵ_0 . In the next section, we shall see that this result is general.

1.14 Discontinuities of the Electric Field

In the last section, we found, in several simple cases, that the perpendicular component of the field has a discontinuity of σ/ϵ_0 across a sheet of surface density σ . We generalize now the result proving that: *when crossing any surface charge distribution with surface density σ , the component of the field perpendicular to the surface has a discontinuity σ/ϵ_0 , while the tangential component is continuous.* We shall meet such charge distributions on the surfaces of the conductors and of the dielectrics.

Consider the oriented surface Σ shown in Fig. 1.24. Let \mathbf{n} be its positive normal unit vector and P_1 and P_2 two points very close to the surface and to one another, one on the negative side and one on the positive. Let E_n be the perpendicular component of the field, namely its component on \mathbf{n} . Figure 1.24 shows the section of a small box having bases at P_1 and P_2 , both of infinitesimal area dA and lateral surface infinitesimal of an superior order to that of dA . We apply the Gauss law to the box. Whatever the components of the electric field can be, the flux through the lateral surface is infinitesimal of superior order by construction and can be neglected. Taking into account that E_n has the direction of \mathbf{n} on one side and the opposite on the other, the Gauss law gives $[E_n(P_2) - E_n(P_1)]dA = \sigma dA/\epsilon_0$.

Hence, the *perpendicular component discontinuity* of the field, say $\Delta E_n \equiv E_n(P_2) - E_n(P_1)$, is

$$\Delta E_n = \frac{\sigma}{\epsilon_0}. \quad (1.81)$$

We now consider the components parallel to the charged surface. We used the plural because we have two of them. Indeed, let us consider a plane tangent to Σ at the point we are considering. The projection on this plane (and locally on Σ) is a

Fig. 1.24 A surface charge distribution and a Gaussian surface

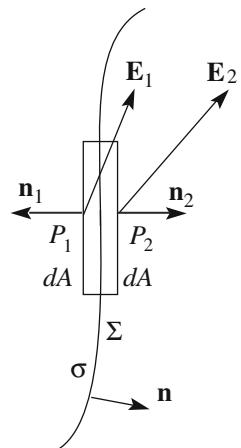
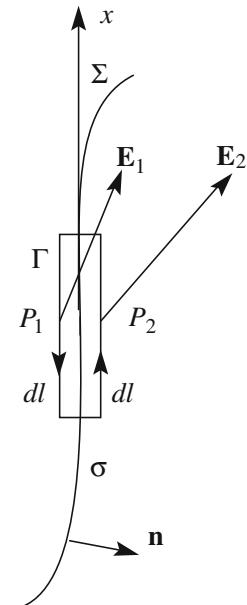


Fig. 1.25 A surface charge distribution and a closed path integral



two-dimensional vector that we call \mathbf{E}_t , whose components are both tangent to Σ . To study their behavior, we shall now exploit the fact that the field is conservative. Consider again two points P_1 and P_2 very close to the surface and to one another, one on the negative side and one on the positive. Let us take the origin of our reference system at the point of Σ between P_1 and P_2 , the z -axis normal to Σ in the direction from P_1 to P_2 and x and y on the plane tangent to Σ . Figure 1.25 is drawn on the xz plane. Consider the x -components of the field, say \mathbf{E}_{tx} , on both sides of Σ .

Figure 1.25 looks similar to Fig. 1.24, but there is a difference, namely the small rectangle here represents a curve Γ . Its longer sides are parallel to Σ through P_1 and P_2 and have equal infinitesimal lengths dl . The other two sides, normal to Σ , are infinitesimal of an order superior to that of dl . We now impose that the circulation of the field about Γ be zero. The contributions along the perpendicular sides are negligible by construction. Taking into account that the direction of \mathbf{E}_{tx} is equal to that of Γ on one side, and opposite on the other, we can write $[E_{tx}(P_2) - E_{tx}(P_1)]dl = 0$, namely the x -component of the field is continuous. The same argument holds for its y -component. Hence, the *tangential component* of the field is continuous, namely the difference between the two sides $\Delta\mathbf{E}_t \equiv \mathbf{E}_t(P_2) - \mathbf{E}_t(P_1)$ is

$$\Delta\mathbf{E}_t = 0. \quad (1.82)$$

1.15 Poisson and Laplace Equations

Let us recall the two fundamental properties of an electric field in a static condition: it is conservative and obeys the Gauss law. The expressions in differential form are

$$\nabla \times \mathbf{E} = 0 \quad (1.83)$$

and

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}. \quad (1.84)$$

We start by recalling Eq. (1.39), which is equivalent to Eq. (1.83), namely

$$-\nabla \phi = \mathbf{E}. \quad (1.85)$$

We substitute this equation in Eq. (1.84), obtaining $\nabla \cdot \nabla \phi = -\rho/\epsilon_0$. On the left-hand side of this expression, we have the divergence of the gradient of a scalar function. This is called the *Laplacian*, and in Cartesian coordinates, it is the sum of the pure second partial derivatives with respect to the coordinates, namely the square of the gradient, which is a vector operator, namely

$$\nabla \cdot \nabla = \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \quad (1.86)$$

The differential equation in ϕ is, in conclusion,

$$\nabla^2 \phi = -\frac{\rho}{\epsilon_0}. \quad (1.87)$$

This is an inhomogeneous second order partial differential equation in ϕ that is explicitly written as

$$\frac{\partial^2 \phi(x, y, z)}{\partial x^2} + \frac{\partial^2 \phi(x, y, z)}{\partial y^2} + \frac{\partial^2 \phi(x, y, z)}{\partial z^2} = -\frac{\rho(x, y, z)}{\epsilon_0}. \quad (1.88)$$

This is a famous equation in mathematical physics called the *Poisson equation*. Solving the equation, one obtains the potential and, hence, the electric field, everywhere. As a matter of fact, we already know the expression of the solution. This is just the potential of the given charge distribution Eq. (1.37), namely

$$\phi(\mathbf{r}_1) = \frac{1}{4\pi\epsilon_0} \int_V \frac{\rho(x_2, y_2, z_2)}{r_{21}} dV_2 + \text{const} \quad (1.89)$$

where V is the volume containing the charge. Mathematics rigorously proves that this is the solution provided that, at infinity, the function ϕ vanishes at least as $1/r$ and its gradient at least as $1/r^2$. In practice, this means requesting that there are no charges at infinite distance.

In a region of space deprived of electric charges, namely where $\rho = 0$, the Poisson equation becomes

$$\nabla^2 \phi = 0. \quad (1.90)$$

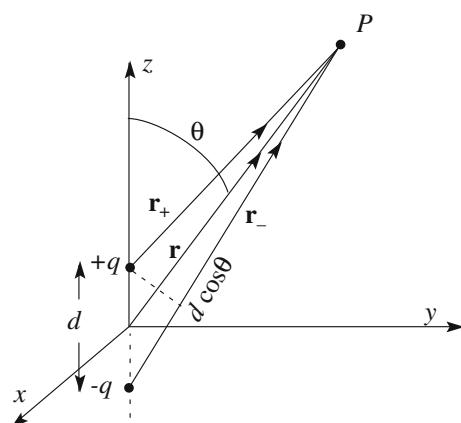
This homogeneous equation is known as the *Laplace equation*. The equation has been the object of extensive study by mathematicians and is very important for physics as well. A function that is a solution of the Laplace equation is called a *harmonic function*. We shall use the Laplace equation in the subsequent chapters.

In this section, we met two fundamental equations of mathematical physics, which are named after two of the greatest mathematicians of all times, Pierre Simon de Laplace (France, 1749–1827) and Siméon Denis Poisson (France, 1781–1840)

1.16 The Electric Dipole

The simplest charge distribution beyond the single point charge, sometimes called a monopole, is the dipole. It consists of two equal and opposite charges, say $+q$ and $-q$, at a fixed distance d . Let \mathbf{d} be the vector from $-q$ to $+q$ and let us choose a reference frame with the z -axis in the positive direction of \mathbf{d} and the origin at the middle point of the dipole. Let P be the generic point where we want to express the potential, \mathbf{r} its position vector from the origin and \mathbf{r}_+ and \mathbf{r}_- the vectors from the positive and negative charge to P , respectively, as shown in Fig. 1.26.

Fig. 1.26 A dipole and a reference frame



The potential in P is

$$\phi(P) = \frac{1}{4\pi\epsilon_0} \left(\frac{q}{r_+} - \frac{q}{r_-} \right) = \frac{q}{4\pi\epsilon_0} \frac{r_- - r_+}{r_- r_+}.$$

This expression is exact. We now look for an approximation valid at distances from the dipole much larger than its size, namely for $r_- \gg d$ (and, consequently, also $r_+ \gg d$). Under these conditions, we can write, approximately, $r_- r_+ \approx r^2$ and $r_+ - r_- \approx d \cos \theta$, where θ is the polar angle, namely the angle between \mathbf{d} and \mathbf{r} , and we get

$$\phi(P) = \frac{q}{4\pi\epsilon_0} \frac{d \cos \theta}{r^2}.$$

We define the vector quantity $\mathbf{p} = q \mathbf{d}$ as the *electric dipole moment*, and we write the last equation, namely its potential, at distances much larger than the dipole size as

$$\phi(P) = \frac{1}{4\pi\epsilon_0} \frac{\mathbf{p} \cdot \mathbf{r}}{r^3} = \frac{1}{4\pi\epsilon_0} \frac{\mathbf{p} \cdot \mathbf{u}_r}{r^2}. \quad (1.91)$$

Note the following. First, the dipole potential does not depend separately on d and q , but only on the dipole moment. Secondly, the potential decreases as the inverse distance square, rather than as the inverse distance, as with the monopole.

QUESTION Q 1.4. An electron and a proton are separated by a 100 pm distance. What is the electric dipole moment of the system? \square

The electric field is the gradient of the potential. Let us start from its axial component z . We have

$$E_z = -\frac{\partial \phi}{\partial z} = -\frac{p}{4\pi\epsilon_0} \frac{\partial}{\partial z} \left(\frac{z}{r^3} \right) = -\frac{p}{4\pi\epsilon_0} \left(\frac{1}{r^3} - \frac{3z^2}{r^5} \right) = \frac{p}{4\pi\epsilon_0} \frac{3 \cos \theta - 1}{r^3}. \quad (1.92)$$

Similarly, the x and y components are

$$E_x = \frac{p}{4\pi\epsilon_0} \frac{3xz}{r^5}; \quad E_y = \frac{p}{4\pi\epsilon_0} \frac{3yz}{r^5}. \quad (1.93)$$

The cylindrical symmetry of the system suggests that we look for the perpendicular field component, namely the projection of \mathbf{E} in the xy plane. This is

$$E_{\perp} = \frac{p}{4\pi\epsilon_0} \sqrt{E_x^2 + E_y^2} = \frac{p}{4\pi\epsilon_0} \frac{3 \cos \theta \sin \theta}{r^3}. \quad (1.94)$$

Fig. 1.27 The electric dipole field force lines

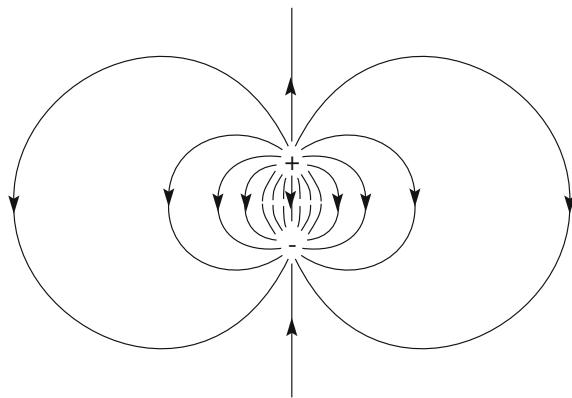
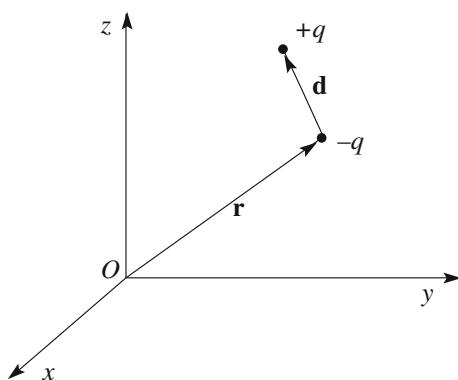


Fig. 1.28 A dipole in an external field



Notice, from Eqs. (1.92) and (1.94), that the dipole field decreases with the inverse of the cube of the distance. The field lines are shown in Fig. 1.27. At a given distance, the field is stronger in the polar direction than in the equatorial. On the axis, outside the dipole, the field has the positive direction of the dipole, and the opposite at the equator. The field intensity on the axis is twice that at the equator at the same distance.

Consider now a dipole in an external electric field. Let ϕ be the external electrostatic potential. We take the reference frame as shown in Fig. 1.28. Let \mathbf{r} be the position vector of the negative charge. The position vector of the positive charge is then $\mathbf{r} + \mathbf{d}$. The dipole potential energy is $U = q\phi(\mathbf{r} + \mathbf{d}) - q\phi(\mathbf{r})$. Consider now a situation in which d is small enough compared to the distance over which the potential varies sensibly. Let $\partial\phi/\partial d$ be the directional derivative of ϕ in the direction of \mathbf{d} . We have

$$\phi(\mathbf{r} + \mathbf{d}) - \phi(\mathbf{r}) \cong \frac{\partial\phi}{\partial d} d = \nabla\phi \cdot \mathbf{d} = -\mathbf{E} \cdot \mathbf{d}.$$

In conclusion, we write

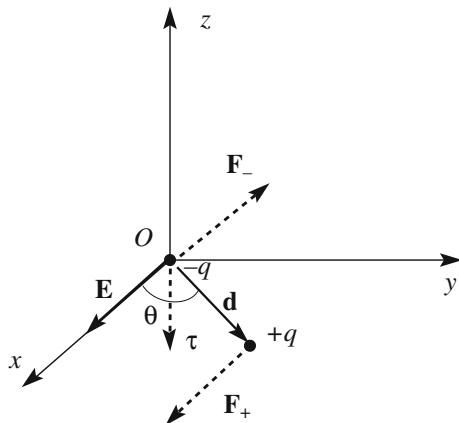
$$U = -\mathbf{E} \cdot \mathbf{p}. \quad (1.95)$$

Notice the following. First, the dipole potential energy in an external field, in the hypotheses we considered, does not depend separately on q and d but only on the dipole moment. Secondly, we have not specified the point of the dipole at which the electric field is to be taken. This is irrelevant in the assumption we made having very small variations of the potential, and consequently of the field too, over the dipole.

Consider now a dipole constrained at its middle point, but free to rotate about it. Looking at Eq. (1.95), we see that the system has two equilibrium positions. One of these is stable, corresponding to the minimum of the potential energy. This is when \mathbf{p} has the same direction and sense as \mathbf{E} . The other position is unstable, being at a maximum of potential energy. This is when \mathbf{p} is parallel and opposite to \mathbf{E} . To better understand the situation, think about the forces acting on the two charges. In both cases, they form a zero-arm couple, but in the former case, they pull the ends of the dipole (stable equilibrium), while in the second, they push them (unstable equilibrium).

Consider now the dipole in a generic position. Let θ be the angle made by its moment with the external field, as shown in Fig. 1.29. In the approximation we made, the field can be considered uniform over the dipole. Consequently, the resultant force acting on its charges is zero (we shall see instances later for which this is not the case). Let τ be the torque on the dipole, which we now wish to express. To do that, we choose a reference with the origin in the negative charge, the x -axis in the direction of the external electric field \mathbf{E} and the y -axis in the plane of \mathbf{E} and \mathbf{p} . Let θ be the angle leading from \mathbf{E} to \mathbf{p} in the direction seen as counterclockwise from the z -axis, as in Fig. 1.29.

Fig. 1.29 Forces acting on a dipole in an external field



Let us apply the virtual works principle. The work done for a virtual rotation by $d\theta$ is equal to the opposite of the corresponding change of potential energy, namely $\tau_z d\theta = -dU$, or

$$\tau_z = -\frac{dU}{d\theta} = -\frac{d}{d\theta}(-pE \cos \theta) = -pE \sin \theta$$

The minus sign on the right-hand side means that the moment τ tends to move \mathbf{p} parallel to and in the positive sense of \mathbf{E} . It is immediate to see that τ is the (axial) vector

$$\tau = \mathbf{p} \times \mathbf{E}. \quad (1.96)$$

If the assumption made so far about the uniformity of the external field over the dimensions of the dipole does not hold, the resultant external force on the dipole is not zero. Let \mathbf{F} be that resultant. We find it using, again, the virtual works principle. Let us choose a direction s and look for the component F_s of \mathbf{F} in that direction. We must think of the dipole as only being free to move in that direction, namely we must consider the virtual work for an infinitesimal rigid translation in that direction. The corresponding energy variation is $dU = -\mathbf{F} \cdot d\mathbf{s} = -F_s ds$. In particular, for the virtual displacements in the directions of the Cartesian axes, we have

$$F_x = -\frac{\partial U}{\partial x}, \quad F_y = -\frac{\partial U}{\partial y}, \quad F_z = -\frac{\partial U}{\partial z}.$$

We take the partial derivatives of Eq. (1.95), starting with x , remembering that the vector \mathbf{p} is constant for the translations we are considering, and obtain

$$F_x = -\frac{\partial U}{\partial x} = -\frac{\partial}{\partial x}(-p_x E_x - p_y E_y - p_z E_z) = p_x \frac{\partial E_x}{\partial x} + p_y \frac{\partial E_y}{\partial x} + p_z \frac{\partial E_z}{\partial x}.$$

We can express the right-hand side in terms of the sole x -component of the field with the following argument. We write, for example, for the y -component,

$$\frac{\partial E_y}{\partial x} = -\frac{\partial^2 \phi}{\partial x \partial y} = -\frac{\partial^2 \phi}{\partial y \partial x} = \frac{\partial E_x}{\partial y}.$$

Doing that, we get

$$F_x = p_x \frac{\partial E_x}{\partial x} + p_y \frac{\partial E_x}{\partial y} + p_z \frac{\partial E_x}{\partial z} = \mathbf{p} \cdot \nabla E_x$$

Finally, the resultant force on the dipole is

$$\mathbf{F} = (\mathbf{p} \cdot \nabla E_x, \mathbf{p} \cdot \nabla E_y, \mathbf{p} \cdot \nabla E_z) = (\mathbf{p} \cdot \nabla) \mathbf{E}. \quad (1.97)$$

We see that the resultant force is proportional to the gradients of the components of the electric field.

1.17 Dipole Approximation

Consider now a distribution of charges at rest limited in a certain arbitrary volume. We shall now find a very useful approximate expression for the potential, and, hence, for the field, at a point P that, compared to its diameter, is at a large distance from the distribution. To be concrete, we consider point charges q_i . Our reasoning can be trivially extended to a continuous distribution.

Let O be the origin of the reference frame at an arbitrary point inside the volume of the distribution and \mathbf{d}_i the position vector of the generic charge q_i . Let \mathbf{R} be the position vector of point P where we want the potential and \mathbf{r}_i the vector from q_i to P , as shown in Fig. 1.30. Obviously, it is $\mathbf{r}_i = \mathbf{R} - \mathbf{d}_i$. The potential at P is

$$\phi(P) = \frac{1}{4\pi\epsilon_0} \sum_i \frac{q_i}{r_i}, \quad (1.98)$$

which is exact. However, in the hypothesis we made, we have $R \gg d_i$ for all i . In a very first approximation (if P is very far away), we can simply consider all the r_i to be equal to R , obtaining

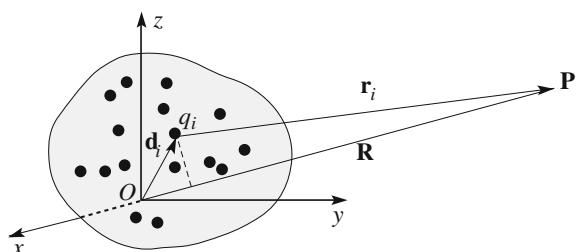
$$\phi(P) = \frac{1}{4\pi\epsilon_0} \frac{1}{R} \sum_i q_i = \frac{1}{4\pi\epsilon_0} \frac{Q}{R} \quad (1.99)$$

where Q is the total charge

$$Q = \sum_i q_i. \quad (1.100)$$

This simply means that, at a large distance, the distribution appears as point-like.

Fig. 1.30 A charge distribution seen from a distance



That which we have just considered is called monopole approximation and is often insufficient. It is certainly so when the total charge is zero ($Q = 0$). This is a very common situation. For example, atoms and molecules are globally neutral charge distributions. Let us then consider the next order approximation, which is the dipole approximation.

Looking at Fig. 1.30 and thinking that P is at a large distance, we understand that the projection of \mathbf{d}_i on \mathbf{R} is approximately equal to the difference between R and r_i . Hence, calling \mathbf{u}_R the unit vector of \mathbf{R} , we can write $r_i \cong R - \mathbf{d}_i \cdot \mathbf{u}_R$. In Eq. (1.98), we have its reciprocal, which can be approximated as

$$\frac{1}{r_i} \cong \frac{1}{R(1 - \frac{\mathbf{d}_i}{R} \cdot \mathbf{u}_R)} \cong \frac{1}{R} \left(1 + \frac{\mathbf{d}_i}{R} \cdot \mathbf{u}_R \right).$$

We substitute in Eq. (1.98), obtaining

$$\phi(P) = \frac{1}{4\pi\epsilon_0} \left(\frac{Q}{R} + \frac{\mathbf{u}_R}{R^2} \cdot \sum_i q_i \mathbf{d}_i \right). \quad (1.101)$$

This expression contains the first two terms of a series expansion. The first term is the monopole that we already found and is clearly the most important at large R when $Q \neq 0$. The second term decreases faster, namely as $1/R^2$, but is the dominant one if $Q = 0$. It can immediately be recognized as a dipole potential, by defining as the electric dipole moment of the charge distribution the vector

$$\mathbf{p} = \sum_i q_i \mathbf{d}_i. \quad (1.102)$$

The potential in P when $Q = 0$ is, from Eq. (1.101),

$$\phi(P) = \frac{q}{4\pi\epsilon_0} \frac{\mathbf{p} \cdot \mathbf{u}_R}{R^2}. \quad (1.103)$$

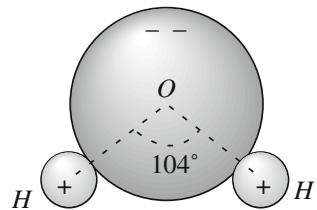
The definition of Eq. (1.102) generalizes the expression $\mathbf{p} = q \mathbf{d}$ valid for a dipole, to which it reduces, as is immediately verified, for two equal and opposite charges.

The definition of electric dipole moment Eq. (1.102) apparently depends on the choice of the origin O of the reference frame, because the vectors \mathbf{d}_i depend on this choice. We now show that this is true if $Q \neq 0$, but for a globally neutral distribution, \mathbf{p} is independent of this choice. Indeed, let us take as the origin another point, say O' . The moment becomes

$$\mathbf{p}' = \sum_i q_i \mathbf{d}'_i = \sum_i q_i (OO' + \mathbf{d}_i) = \sum_i q_i \mathbf{d}_i + OO' \sum_i q_i = \mathbf{p} + OO' Q.$$

Hence, \mathbf{p}' is equal to \mathbf{p} if Q is zero.

Fig. 1.31 Schematic representation of a water molecule



Equation (1.103) shows that the potential, and hence the electric field, at a large distance from a globally neutral charge distribution is equal to that of a dipole (two equal and opposite charges) with the same moment. Let us define more precisely such an equivalent dipole. Let us mentally separate the positive charges, which we call q_i^+ , and the negative ones, which we call q_i^- . Let Q^+ and Q^- be their sums, respectively ($Q^- = -Q^+$). We define as the centers of the positive and negative charges, in analogy with the center of mass of a mass distribution, the points identified by the position vectors

$$\mathbf{d}^+ = \frac{\sum_i |q_i^+| \mathbf{d}_i}{\sum_i |q_i^+|} = \frac{\sum_i |q_i^+| \mathbf{d}_i}{Q^+}, \quad \mathbf{d}^- = \frac{\sum_i |q_i^-| \mathbf{d}_i}{\sum_i |q_i^-|} = \frac{\sum_i |q_i^-| \mathbf{d}_i}{|Q^-|}. \quad (1.104)$$

We can now write Eq. (1.102) as

$$\mathbf{p} = \sum_i |q_i^+| \mathbf{d}_i^+ - \sum_i |q_i^-| \mathbf{d}_i^-,$$

which, for Eq. (1.104), gives us

$$\mathbf{p} = Q^+ (\mathbf{d}^+ - \mathbf{d}^-) = Q^+ \mathbf{d}, \quad (1.105)$$

having defined the vector from the center of the negative charge to the center of the positive charge as $\mathbf{d} = \mathbf{d}^+ - \mathbf{d}^-$.

In conclusion, the electric field of a globally neutral charge distribution at distances that are large compared to its size is equal to the field of an electric dipole with all the negative and positive charges concentrated at the respective centers. As a consequence, the dipole moment is zero when the two centers coincide. This is the case for atoms and molecules with spherical symmetry. They do not have an “intrinsic” electric dipole moment. However, very often, molecules do have an electric dipole moment. A very important example is the water molecule. Its rather large dipole moment has a strong influence on the physics of water.

Figure 1.31 gives a schematic representation of the H_2O molecule. The oxygen atom has a negative net charge, while the hydrogen atoms have net positive charges. From the figure, it is clear that it is not a dipole. However, the field of the molecule is the dipole field at distances that are large compared to its diameter, which is about 0.29 nm.

The dipole moments of an unperturbed molecule are called *intrinsic electric dipole moments*. As we shall see in Chap. 4, external electric fields acting on a molecule modify its charge distribution, producing an induced moment even on the molecules with zero intrinsic moment.

Summary

In this chapter, we have learnt the following principal concepts:

1. The constituents of matter are charged with positive and negative elementary charges. The elementary negative charge is exactly equal to and opposite of the positive one.
2. The force between two point charges at rest in a vacuum is inversely proportional to the square of their distance.
3. The concept of (static) electric field.
4. The method for calculating the electrostatic field produced by a known distribution of charges in fixed positions in a vacuum.
5. That the electrostatic field is conservative (or irrotational) and the concept of electrostatic potential.
6. The method for calculating the electrostatic potential produced by a known distribution of charges in fixed positions in a vacuum.
7. The measurement of the elementary charge.
8. The flux of the electrostatic field and its properties.
9. The equations for the divergence and the curl of the electrostatic field.
10. The Poisson and Laplace equations.
11. The discontinuities of the electrostatic field.
12. The electric dipole and the dipole approximation of the field of a globally neutral charge distribution.

Problems

- 1.1. Does a closed line of force exist in an electrostatic field? Do lines of force irradiating from a point exist?
- 1.2. How do you calculate the electric field and the potential of three point charges q_1 , q_2 and q_3 , at rest, respectively, at \mathbf{r}_1 , \mathbf{r}_2 and \mathbf{r}_3 ?
- 1.3. Can one produce an electric field of 1 kV/m having a battery of 1.5 V?
- 1.4. Describe a measurement of the elementary charge.
- 1.5. What is the relative precision needed to show the discrete nature of charge measuring a 1 nC charge?

- 1.6. Let S_1 and S_2 be two closed surfaces, with S_1 being completely contained in S_2 . The flux outgoing from S_1 is positive. The flux outgoing from S_2 is zero. What can you say about the charges generating the field?
- 1.7. A proton moves toward a potassium nucleus ($Z = 19$). Find the repulsive electrostatic force when it is at 50 fm.
- 1.8. A spherical soap bubble has a charge of $Q = 100 \text{ pC}$. What are the fields in its center and at a point immediately outside? Are all the components of the field continuous crossing its surface?
- 1.9. A charge Q is uniformly distributed in the volume of a sphere of radius R . Find the field and the potential at its center.
- 1.10. Consider the electric field $\mathbf{E} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$, where a, b, c are constants and $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are the unit vectors of the reference frame. Is the field uniform? What is the expression of the potential?
- 1.11. The electric dipole moment of a water molecule is $p = 6.1 \times 10^{-30} \text{ C m}$. How far should be a proton be from an electron for them to have the same dipole moment? Compare the result with the water molecule diameter. Calculate the electric field in the equatorial plane of the dipole at 1 nm distance from axis.
- 1.12. Find the electric field at the distance of 0.5 nm from a monovalent ion (taken as point-like).
- 1.13. What are the properties of the electric dipole moment of a globally neutral charge distribution?
- 1.14. A charged droplet of mass $m = 1 \text{ fg}$ is between the parallel and horizontal plates of a capacitor (using to produce a uniform vertical electric field). When no field is present, the droplet falls at a certain constant speed. If a field of $3 \times 10^5 \text{ V/m}$ is applied, the droplet falls at half that velocity. How much is the charge on the droplet? To how many elementary charges does it correspond?
- 1.15. An electron at rest at 10 nm distance from a water molecule is subject to a certain electric force. How does this force vary if the distance increases to 20 nm? (Assume that the shape of the molecule does not vary.) Compare the result with the force between two elementary charges at the same distance.
- 1.16. Two horizontal plane surfaces are uniformly charged. The charge densities are $\sigma_1 = +3 \text{ nC m}^{-2}$ on the lower plane and $\sigma_2 = -6 \text{ nC m}^{-2}$ on the upper one. What is the electric field between the planes and in the spaces above and below them?

Chapter 2

Conductors in Equilibrium

Abstract Conductors are characterized by their containment of electric charges, electrons or ions, which are free to move about inside their body. These are the metals, the electrolytes and the ionized gases. In this chapter, we study the behavior of the conductors under electrostatic conditions, namely constant in time. We shall study electrostatic induction between two and more conductors, the capacitors and the electrostatic shield.

In the first chapter, we studied the electrostatic field in a vacuum generated by charges in fixed positions. In real life, material bodies are always present, containing an enormous number of electric charges. Consequently, under practical circumstances, the electric field is due both to the charges under our control and to the charges of the surrounding bodies. The building blocks of matter, molecules and atoms, contain positive and negative charges in equal quantities and are consequently globally neutral. The electric field generated by these nanoscopic structures is zero at distances that are large compared to their size under unperturbed conditions. Contrarily, an externally applied field may deform and move the molecules, as well as the free charges existing in the conductors, changing the internal charge distribution and contributing the macroscopic field.

The electric field we shall consider is “macroscopic”, because our description shall always deal with macroscopic phenomena, which happen on distance scales that are enormous compared to the sub-nanometer scale of the molecules and on time scales quite significantly longer than the characteristic times of atomic phenomena. Consider, for example, a piece of metal between the plates of a capacitor. The field generated by the capacitor is a macroscopic field. If we look into the metal at the atomic scale, we shall see a field that varies very rapidly from point to point; it is very strong between a nucleus and the electrons, but it is almost zero outside the atom. In addition, the field is far from being constant, but varies over very short times depending on the motions of the electrons. At the atomic level, there is never

an electrostatic field, but rather an electric field that rapidly varies in time and with position. This is an extremely complicated situation, but it is not what we are interested in. We are interested in the field we can measure with our instruments. These instruments are macroscopic bodies with geometrical dimensions much larger than the molecular ones that integrate on time scales much longer than those of the microscopic variations. This is the macroscopic field.

The material bodies can be schematically divided into two classes; the conductors, studied in this chapter, and the dielectrics or insulators, studied in Chap. 4. In this chapter, we shall always deal with static, namely time independent, conditions.

In Sect. 2.1, we shall see how the conductors containing elementary charges can move freely. These charge carriers, as they are called, are electrons in metals, ions in electrolytes and in gases, “holes” and electrons in semiconductors, etc. Under static conditions, the charge carriers randomly move, similarly to the molecules of a gas. There is no ordered motion under these conditions.

In Sect. 2.2, we shall see that when a conductor is brought into an external electric field, some of its free charges move to the surface and arrange themselves in such a way as to cancel, with the field they produce, the applied field in the entire volume of the conductor. Under static conditions, the electric field is zero in a conductor, which is consequently an equipotential volume. The charge density inside the conductor is also zero, while a surface charge density can be present on the surface.

Suppose we want to have an electric field of certain intensity and certain shape in a region of space. The way to produce it is to build a number of conductors having surfaces of the right shape and to give them the right potentials. In doing that, we control the potentials, not the charges. The underlying mathematical problem is finding the potential in a region of space once the values on the surfaces surrounding the region are fixed. We shall see in Sect. 2.4 that the solution of the problem is unique and we shall study some of its properties. We shall also see that the only available general means to find the solution are not analytical, but numerical.

In Sect. 2.6, we shall consider a system of two conductors and learn the conditions that must be satisfied for it to be a capacitor. Capacitors are important elements of any electronic and electric circuit. We shall study their properties in Sects. 2.7 and 2.8. In Sect. 2.9, we shall extend the study to systems of more than two conductors.

In Sect. 2.10, we shall study the electric field of a system of conductors completely surrounded by a hollow conductor. We shall see how the latter divides the space into two independent regions, as far as electrostatic phenomena are concerned. This is the electrostatic shielding action. This action is a consequence of the inverse square law of the electrostatic force and its study allows for verifying this dependence with extreme precision. We shall give examples of the practical importance of electrostatic shielding.

2.1 Conductors

A body is defined as an electric conductor if it contains, in its bulk or on its surface, electric charge carriers that are free to move. As we know, all bodies contain, or, to be more precise, are made of, charged particles, such as electrons and nuclei. These charged particles are inside atoms or molecules, which are globally neutral and are generally not free to move over macroscopic distances. In the conductors, a small fraction of these charges can move over macroscopic distances. The principal types of conductors are as follows.

Metals.

Metals in their solid phase are made of microcrystals. A small fraction of the electrons, typically one or two per atom, are free to move about inside the metal. These are called free electrons. To fix the orders of magnitude, typical free electron densities in metals are $n_p = 10^{29}\text{--}10^{30}/\text{m}^3$. With a good approximation, at the usual temperatures, we can think of the free electrons as behaving like a gas, with a velocity distribution similar to that of the molecules of a common gas. Their mean kinetic energy at room temperature is like that of a gas, in a round figure, 1/40 eV. The electrons are free inside the metal, but cannot leave it because they are attracted by the array of positive ions that they have abandoned. The energy needed to be given to an electron to extract it from the metal depends on the metal, but is generally of a few electronvolt, much larger than their mean kinetic energy at room temperature.

Semiconductors.

Different types of semiconductor exist, both natural and artificially produced. We shall only mention the simple example of a pure element, the Si. The Si atom has four valence electrons, namely four electrons taking part in its chemical bonds. In a Si crystal, every atom is linked to four other atoms through covalent bonds. A covalent bond is made of two electrons, one for each partner atom. The bonds are very stable and very few of them break, freeing electrons. Technologies are available to dope the crystal, namely to grow it in the presence of a few impurities.

Doping can be done with a pentavalent element, the atoms of which are fit to substitute for Si atoms in the crystal lattice. These extraneous atoms do form four links with four Si atoms, but are left with an extra valence electron. This electron is only weakly bound and detaches due to the thermic motion already at low temperatures. This is the case of the n-type semiconductors, where “n” stands for negative, because the charge carriers are the electrons. The carrier density depends on the applied doping level, but is always much smaller than in metals, because only a small number of Si atoms is substituted by the dopant. A typical order of magnitude is $n_p = 10^{20}/\text{m}^3$.

If Si is doped with a trivalent element, we obtain a p-type semiconductor, where “p” stands for positive, because such is the charge of the carriers. When a trivalent atom substitutes for a Si in the ladder, it can only establish three bonds. The fourth

is made by one electron alone. There is a “hole”, namely a point where an electron is missing. This situation is not stable, and it soon happens that an electron of a nearby bond jumps into the hole. But, in so doing, another hole is produced and the process repeats and propagates. We can also say that the hole moves in the opposite direction from one position to the other and that a positive charge moves with it. Indeed, holes behave like positively charged particles, with a chaotic motion similar to that of electron gas.

Electrolytes.

Ionic molecules, such as the common salt NaCl, are made by two oppositely charged ions, Na^+ and Cl^- in the example. As we shall see in Chap. 4, the electrostatic force is much weaker in water than in a vacuum. Consequently, the ion bond breaks in water and the two ions become free. A small quantity of ions (of both signs) is always present in water, which is consequently a conductor. The ion carrier density can be reduced by distillation or increased by adding salts.

Gases.

A small fraction of ionized molecules is also always present in gases. Ionization is mainly caused by natural radioactivity from the rocks and from cosmic rays. The charge carriers are ions of both signs.

Other types of conductors are plasmas, some non-metallic substances like graphite, some organic substances, etc.

2.2 Conductors in Equilibrium

In this chapter, we study the properties of the conductors in equilibrium, namely under static conditions. We have already encountered the concepts of charge density and electric field in a vacuum. We now need these concepts in a conductor, which is a material medium. Matter is made of molecules, and inside molecules, there are electrons separated from the nuclei of their atoms by empty spaces much larger than the nuclear diameters. These particles are not at all at rest, but move continuously at high speeds. The charge density that we shall consider is an average taken on “physically infinitesimal” volumes, namely very small relative to the macroscopic dimensions but containing a large number of molecules. This is possible because the atomic radii are on the order of a tenth of a nanometer.

Inside the material, the electric field changes by large factors over distances on the order of atomic diameters. Fields are very intense inside an atom, becoming almost zero immediately outside of it. In addition, at the atomic scale, the field varies very rapidly, namely over times on the order of the femtosecond, due to the fast motion of the electrons. This electric field at the nanometer and femtosecond scales is not the field we measure with our macroscopic instruments, as we already stated in Sect. 1.3 when we gave the operational definition of the field. The

macroscopic field we shall deal with at point \mathbf{r} at the instant t is an average on a volume around \mathbf{r} , very small in macroscopic dimensions but very large on the nanometric scale, and on a time interval around t , small on the macroscopic scale but very long on the femtosecond scale.

The conductors' properties we shall now discuss only hold for homogeneous conductors. This means that temperature, chemical composition, aggregation phase (in short, all their chemical and physical properties) are independent of position.

We shall now discuss four properties. They are very simple, but there is a logical order (that should be remembered) in which they have to be considered.

1. *Within a conductor in equilibrium, the charge density is zero.*

Indeed, if the field is not zero, there are forces acting on the free charges. These would accelerate and the condition would not be static. Let us look more closely at what happens.

Consider, for example, a metal, and let us bring it into an electric field. Initially, the field penetrates inside the metal and exerts forces, in particular, on the free electrons. Being unbound, they start moving in the direction opposite to the field. In this way, two charge accumulations develop, a negative one on the side to which electrons move, and a positive one on the opposite side. To be clear, we anticipate that these accumulations are on the surface of the body, as we shall soon see. These charge densities produce an electric field having a direction opposite to that of the external field and tending to cancel it out. As a matter of fact, the free carriers *must* adjust their position to cancel the external field completely. Otherwise, they have not yet reached the equilibrium state. Note that not all the free charges need to move, a very small fraction being sufficient. Note also that the time needed to reach the equilibrium in a metal is extremely short.

We can easily see that this property does not hold for an inhomogeneous conductor. Consider, for example, a metal bar whose extremes are at different temperatures. The conduction electron gas has a higher temperature, hence a higher mean kinetic energy at one extreme than at the other. The carriers will then also move from the hotter to the colder extreme when no external field is present. Under these conditions, the equilibrium is reached when the field in the conductor has a certain, non-zero, value due to the distribution of its carriers.

2. *A conductor is an equipotential volume.*

This is an immediate corollary of the previous property. Indeed, the potential difference between two points of the conductor is the line integral of the field between those points. If the points are inside, we can always find a line completely inside the conductor, on which the field is zero. Hence, all internal points have the same potential. The same is true, by continuity, on the surface. We shall see important applications of this property.

If we want to generate an electric field with a certain shape, we need to fix the potential values in certain space regions. We can do that by building metallic elements of the right shape and giving them the right potentials.

3. The charge density is zero inside a conductor in equilibrium.

Let us apply the Gauss law to an arbitrary closed surface Σ completely inside the conductor. The field is zero at all the points of Σ , and consequently the flux is zero. If ΔV is the volume surrounded by Σ , the Gauss law gives us

$$\int_{\Delta V} \rho dV = 0.$$

ΔV being arbitrary, ρ must be identically zero in the entire inside volume

$$\rho(x, y, z) = 0 \text{ inside.} \quad (2.1)$$

Clearly, this does not mean that there is no charge in the conductor; rather, it means that the net charge is zero.

Note that the just-stated theorem does not hold on the surface. Indeed, a surface element cannot be enclosed in a Gauss surface completely inside the conductor. Indeed, a net free charge exists on the surface in the presence of an external electric field or when the conductor is charged.

We can say that, in general, a charge of density σ , which is a function of the position, is present on the surfaces of the conductors. The integral of σ over the surface is the net charge of the conductor.

4. The electric field immediately outside the surface of a conductor in equilibrium with surface density σ is normal to the surface and has a normal component equal to σ/ϵ_0 .

Indeed, all the field components are zero on the internal face of the surface. Through the surface, the tangent ones are continuous, while the normal one has a discontinuity σ/ϵ_0 . More precisely, if \mathbf{n} is the unit normal outgoing vector, the field on the outside face of the surface is

$$\mathbf{E} = \frac{\sigma}{\epsilon_0} \mathbf{n}. \quad (2.2)$$

Note that the field is directed outward if $\sigma > 0$, and inward if $\sigma < 0$.

We now describe electrostatic induction. The phenomenon, not to be confused with electromagnetic induction, which will be discussed in Chap. 7, can be observed with elementary means. Figure 2.1 represents a metallic conductor on an insulating support. Let us bring a charged object, for example, a glass bar charged by friction (the charge is positive in this example), near to an extreme of the conductor. Under these conditions, some negative free charges move to the region of the conductor nearer to the positively charged body. Correspondingly, a net positive charge density develops on the opposite side. This induced charge rearrangement is such as to cancel the electric field in the entire volume of the conductor. However, the net charge of the conductor remains zero, being insulated.

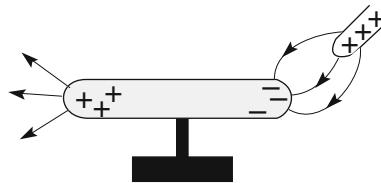


Fig. 2.1 Producing electrostatic induction on an isolated conductor

Consequently, if we now turn away the bar, the positive and negative induced charges in the conductor, so to speak, recombine and the charge density goes back to zero through the entire volume.

A slightly different set-up is shown in Fig. 2.2. Here, the conductor is made of two parts, both on an insulating support. Initially, the two parts are in contact and behave as a single conductor. The effect of the glass charged bar in Fig. 2.2a is equal to that in Fig. 2.1. With the bar still in position, we now separate the two parts, as in Fig. 2.2b, touching the insulating supports (our body is a conductor). When we turn the bar away, the induced negative and positive charges can no longer recombine. In each part, the charge redistributes in order to cancel the field inside the conductor. We end up with two conductors whose surfaces are charged with opposite sign surface charge densities. We say that the bodies have been charged by induction.

The induction phenomenon is easily observed with a gold-leaf electroscope. If we place a body charged by friction, like a glass or plastic bar, near the small sphere of the instrument, without touching it, we observe the leaves opening up, as in Fig. 2.3. If we remove the charged body, the leaves close back up. The phenomenon produced by the charged bar is clearly the electrostatic induction we have just described. If the bar has positive charges, the charges induced on the leaves, which are in the farther side of the electroscope conductor, are positive too, while negative charges are induced on the upper sphere. When we take the charged body away, the induced charges recombine and the leaves discharge.

Figure 2.4 shows how to charge an electroscope permanently by induction. To do that, we just touch the sphere of the electroscope with a finger when the inducing bar is still present. Our body being a conductor, this action electrically connects the sphere with the ground. The charges on the sphere, opposite to those on the inducing body, run away as far as they can, namely to the ground. We now remove

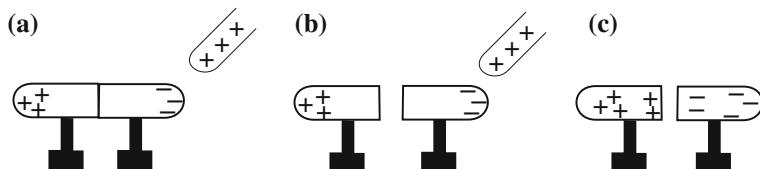


Fig. 2.2 Charging two insulated conductors by induction

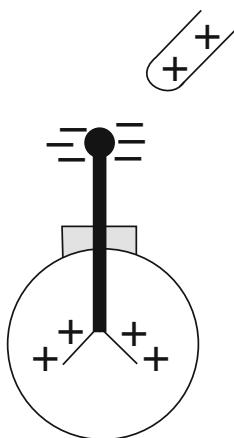


Fig. 2.3 Inducing charges on an electroscope

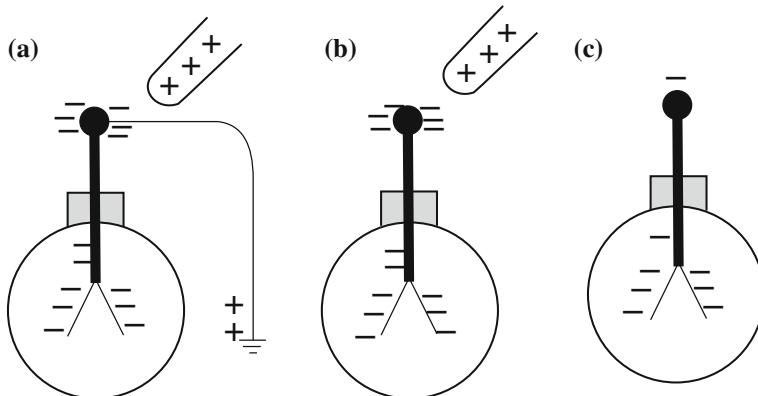


Fig. 2.4 Charging an electroscope by induction

our finger from the sphere and then take the inducing body away. The electroscope is now permanently charged with the sign opposite to that of the inducing body. We can check that by bringing back the charged bar. We see the leaves closing down.

If we observe the leaves of an electroscope we have charged, we see them gradually closing down. The electroscope is discharging because the air in its glass container is not a perfect insulator due to the presence of ions. Ions are continuously formed in air through natural radioactivity. As a matter of fact, we can measure the induced activity rate in air by measuring the discharge time of a properly built electroscope. Here, we open a parenthesis to summarize how cosmic rays were discovered. In 1910, Domenico Pacini (Italy, 1878–1934) developed techniques for measuring the discharge rate of an electroscope underwater. He took measurements three meters underwater in a lake and in the sea, namely under a water thickness

sufficient to absorb the largest fraction of the radiation from the ground. The observed discharge rate had decreased, compared to on shore, but it was still relevant. Pacini concluded that an ionization source different from those in the rocks had to exist. The extraterrestrial origin of the source was established by Victor Hess (Austria, 1883–1964), with a series of balloon ascensions between 1911 and 1912. Hess found that the ionization rate in the atmosphere, as measured by the discharge time of his electroscope, was constant or slightly decreasing up to about 2000 m, somewhat equivalent to Pacini’s 3 m of water. Above 2000 m, the ionization rate monotonically increased up to the maximum altitude of 5000 m that he was able to reach. Hess concluded that the source of the ionizing radiation found by Pacini had an extraterrestrial origin. These are the cosmic rays, as Robert Millikan (USA, 1868–1953) named them, high energy charged particles, protons, nuclei and electrons, coming from the universe.

2.3 Surface Charges on a Conductor

As we have just seen, the electrostatic field just outside the surface of a conductor at a point at which the surface density is σ is normal to the surface and has magnitude σ/ϵ_0 . Consider now the force exerted by the field on the surface charges. Notice first that the force is, in any case, directed outward, namely tending to rip away the charges. Indeed, where σ is positive, the field is directed outwards and the force has its direction, while where σ is negative, the field is inward but the force is opposite to it. As we are dealing with the charge per unit surface, we should consider the force per unit surface as well. This has the dimensions of a pressure and we shall indicate it with P .

One might think this pressure to be the product of the field σ/ϵ_0 and the charge on the unit surface σ , namely σ^2/ϵ_0 , but it is not so. As a matter of fact, the “surface” occupied by the charges is not a geometrical one of zero thickness, but has a finite thickness smaller than the atomic diameters. The electric force acting on the most external layer of charges is, indeed, σ^2/ϵ_0 , but it is zero on the innermost layer, with intermediate values in between. Consequently, we take a mean value, namely one half of the external field. The pressure on the surface charges is then

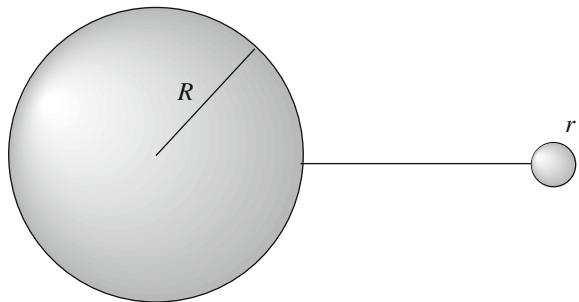
$$P = \frac{\sigma^2}{2\epsilon_0}. \quad (2.3)$$

Note that the pressure might be large enough to pull the charges out. The pressure is higher where the surface curvature is larger, as we shall now see.

Let us start by considering the simple system in Fig. 2.5, consisting of two metal spheres, one larger (radius R) and one smaller (radius r), joined by a metal wire. The system is a single conductor.

Let the system be charged and, at equilibrium, let Q and q be the charges on the large and small spheres, respectively. We want to find the relative values of the two

Fig. 2.5 Two spherical conductors joined by a conductive wire



charges. If the spheres are far enough apart from one another, we can think of the field, or the potential, near one of them as not being affected by the field, or the potential, near the other one. The potential on each surface is then the potential of a charged sphere, which is equal to the potential of a point charge in its center. On the other hand, the two spheres are electrically connected and their potentials must be equal. We can then write $Q/(4\pi\epsilon_0 R) = q/(4\pi\epsilon_0 r)$, or

$$\frac{Q}{R} = \frac{q}{r}. \quad (2.4)$$

The charge on each sphere is proportional to its radius. The field is proportional to the charge density, which is the charge divided by the surface area, that is proportional to the radius squared. Consequently, the charge densities are inversely proportional to the radii, or, we can say, proportional to the curvatures. Calling Σ and σ the charge densities on the large and small spheres, respectively, we have

$$\frac{\Sigma}{\sigma} = \frac{r}{R}. \quad (2.5)$$

Consider now a conductor having regions of different curvature, such as the one shown in Fig. 2.6. In a first approximation, we can consider it similar to the two spheres system and take Eq. (2.5) to be valid near its extremes. The field near the surface, being proportional to the charge density, is proportional to the curvature as well, while the pressure of Eq. (2.3) is proportional to the square of the curvature. Both can be very intense near a tip.

We can verify how the charge is distributed as follows. We use a “spoon” made of an insulating arm finishing in a small conducting sphere. When we touch a point of the conductor with the “spoon”, we take out a charge proportional to the density at that point. We bring this charge to an electroscope and see how much the leaves open. We repeat the operation, taking charges from points of different curvature and verifying that they are larger where the curvature is higher.

Notice that surface curvature may be null or even negative. It is negative in any part of a body folded inward. Where the curvature is negative, the charge density is very small. Consider, for example, a metal conductor having the form of a cone, as in Fig. 2.7. With the method we have just described, it is found that almost no

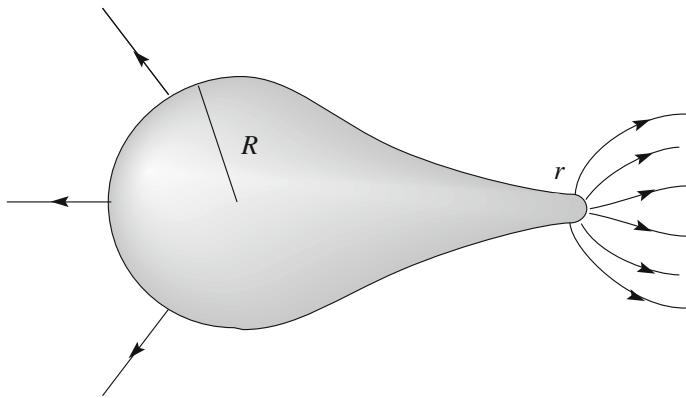


Fig. 2.6 Conductor with different curvatures and lines of field

Fig. 2.7 Demonstration metal cone. The curvature is positive outside, negative inside. Reproduced with permission of the Physics and Astronomy Department of the Padua university



charge is present on the surface inside the cone. We shall see in the next section that the charge density is rigorously zero on the internal surface of a cavity completely enclosed in a conductor.

The above conclusions have relevant practical implications. In several instances, conductors, typically metals, are surrounded by air. If a large potential is given to the conductor, namely if it hosts a large electric charge, and there are points of high curvature (tips) on its surface, the charge density might be extremely high at these points, and so would the electric field just outside the tips. In air, a small number of ions is always present. The ions near the tip are accelerated by the electric field. If the field is high enough, the energy gained by an ion between two collisions with

gas molecules may be so large that, in the next collision, the molecule that is hit gets broken. The two parts of the broken molecule have a charge. Each of them will suffer the same fate as the first ion. An avalanche free charge multiplication process is triggered, air becomes a good conductor and an electric spark suddenly develops. The spark can be very dangerous, both for people nearby and for the equipment. When working with high voltages, any tip must be avoided, and all surfaces must have small curvatures and be smoothed.

A lightning strike is triggered by a similar mechanism. An electric field is always present in the atmosphere; its intensity somewhat varies in time at the ground level, being around 100 V/m on a clear day. During a thunderstorm, the field, between the lower parts of the clouds and the ground, grows to 10^4 V/m, two orders of magnitude larger, over flat surfaces. The field is much higher near high curvature points, like bell towers, trees or even the body of a person, when standing on the flat surface of a beach or on a boat. Field intensity may grow enough so as to trigger an avalanche multiplication process of ions present in the atmosphere. Air becomes a good conductor and an extremely intense current develops between the clouds and the ground, for a short duration. This is the lightning strike. During a thunderstorm, it is imprudent to stay near a tree or other pointed objects, to stand on a beach, or, even more so, to hold pointed conductors in one's hand.

The lightning rod, also called a lightning conductor, was invented in 1749 by Benjamin Franklin (USA, 1706–1790), to protect buildings, ships, etc., in the event of a lightning strike. It is a metal rod mounted on top of a structure for protection, connected to the ground (or sea) by thick copper conductors, capable of “attracting” the spark and discharging the high current into the ground, instead of allowing it to pass through the structure.

2.4 Hollow Conductors

Consider a hollow conductor containing an empty closed cavity. We shall now show that the charge surface density in the cavity is zero. The demonstration has two steps, which should not be inverted. In the first step, we use the Gauss law, in the second, the fact that the electrostatic field is conservative.

It is always possible to find a closed surface enclosing the cavity and entirely inside the conductor, as Σ in Fig. 2.8a. If any charge is present inside Σ , it must be on the surface of the cavity. The flux outgoing from Σ is zero because the field is zero at all of its points. Hence, the total charge on the surface of the cavity is zero. This does not mean that there is no charge but that, if there are charges, there should be as many that are positive as are negative. If there were charges on the surface of the cavity, there would be an electric field, like in the example shown in Fig. 2.8b, in which we have drawn a field line. If this were true, we could always find a closed line like Γ in the figure that follows a field line in the cavity and closes back inside the conductor. The line integral of the field around Γ would certainly be positive, because the field has the direction of the line along the entire part inside the cavity

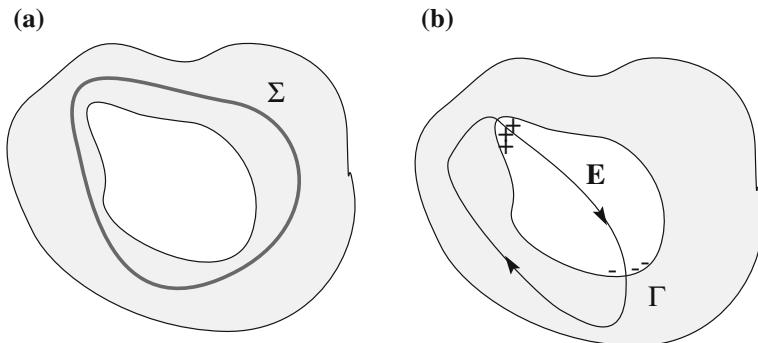


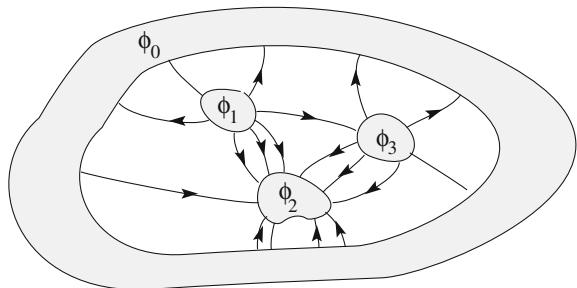
Fig. 2.8 A cavity in a hollow conductor. **a** A surface containing the cavity, **b** a closed path partially in the cavity, partially in the conductor

and is null in the other part. The hypothesis has thus been reduced *ad absurdum* and must be wrong. In conclusion, no charge can be present on the surface of an empty cavity completely enclosed in a conductor in equilibrium. Consequently, the field in the cavity is also zero.

Let us now consider the situation outside the hollow conductor. There, we might have charges of any value in any position outside the conductor and on the conductor itself. The field outside can be anything, while being static. Well, as difficult as it may be to believe, the charges on the external surface of the charged conductor arrange themselves in such a way that no net charge and no field exist, not only in the body of the conductor but even inside the cavity and on its surface. Also, if we put more charge on the conductor of one sign or the other, its potential will rise or fall, but it will never happen, at equilibrium, that any charge or field will appear in the cavity or on its surface. The internal space is completely separated from the external space, for the electrostatic phenomena. This is the electrostatic shielding action. We shall discuss it for cavities that are not empty in Sect. 2.10.

Let us now consider in general terms how one can build an electrostatic field of a desired shape in a given region of space. As we have already mentioned, one can do that by properly shaping a certain number of conductors, metal in general, and giving to each of them a proper potential. Namely, we control the shapes of the equipotential surfaces that are the surfaces of our conductors and their potentials. The charge surface densities, namely how the superficial charges arrange themselves, come about as a consequence. The problem is hence the following: given the surfaces of the conductors and their potentials, find the field and the surface densities. The two quantities are linked. The surface densities are linked to the field intensity on the surface by $\sigma = \epsilon_0 E$. Notice that the problem is different from the problem we discussed in the first chapter of finding the field of a given charge configuration. Now, the charge distribution is not given, as the charges are free to move outside our control, always arranging in such a way as to have a zero field inside the conductors.

Fig. 2.9 Conductor in a closed cavity of a surrounding conductor



Let us analyze the case of n conductors with charges Q_1, Q_2, \dots, Q_n , respectively, enclosed in the cavity of another conductor, which might be charged too, as shown in Fig. 2.9. Notice that this is the situation usually met in practice, when we conduct experiments in a room in a building. Indeed, walls, ceiling and floor behave, in practice, as conductors for static or electric phenomena not too quickly varying in time. Usually their potential is taken as zero (ground).

We start by observing that a charge must exist on the internal surface of the hollow conductor. This charge, which we call Q_0 , is exactly equal and opposite to the sum of the charges of the bodies inside the cavity, namely

$$Q_0 = -Q_1 - Q_2 - \dots - Q_n.$$

To show that, let us consider a surface Σ enclosing the cavity completely inside the conductor. The field is zero at all its points, and consequently the outgoing flux is zero. For the Gauss law, the net charge inside Σ must be zero.

Note that Q_0 is, as we said, the charge on the internal surface of the hollow conductor. It is not its total charge. More charge might be present on the external surface. However, whatever that charge, the charge on the internal surface is always Q_0 .

QUESTION Q 2.1. Two charges Q_1 and Q_2 are at rest at a certain distance. Let F_{12} be the force exerted by Q_1 on Q_2 . You now enclose Q_1 in the center of a spherical metal shell, letting Q_2 be outside. How does F_{12} vary? \square

Let us now analyze the problem more formally. Under the given conditions, the conductors inside have known potentials $\phi_1, \phi_2, \dots, \phi_n$, with the hollow one being ϕ_0 . We are interested in the internal space, in which the Laplace equation holds

$$\nabla^2 \phi = 0 \quad (2.6)$$

with the boundary conditions $\phi(x, y, z) = \phi_1$ at the points of the surface Σ_1 of conductor 1, $\phi(x, y, z) = \phi_2$ at the points of the surface Σ_2 of conductor 2, ..., $\phi(x, y, z) = \phi_0$ at the points of the surface Σ_0 of the hollow conductor. The problem is now precisely posed and we can ask the questions: does a solution exist? If it exists, is it unique? How can we find it? From a physical point of view, it is obvious that the system will reach a unique equilibrium configuration, in which the potential

energy is a minimum. The potentials of the conductors in this configuration are the solution to the problem. This is clearly not a mathematical demonstration. The rigorous demonstration can be found in calculus books. We shall not give it here. Rather, we shall assume that a solution exists and show that it is unique, once the boundary conditions are given.

Let us show a few useful properties of the harmonic functions, as the solution of the Laplace equation are called.

Let V be the space between the conductors and Σ the surface limiting this space, namely the set of the surfaces $\Sigma_0, \Sigma_1, \Sigma_n$. Let us show that if the function ϕ is harmonic in the volume V and is zero at the points of Σ , then ϕ is identically zero in V . We start from the identity valid for every scalar function ϕ

$$\nabla \cdot (\phi \nabla \phi) = \nabla \phi \cdot \nabla \phi + \phi \nabla^2 \phi, \quad (2.7)$$

which is immediately shown by direct calculation (it is mainly the derivative of a product). If ϕ is now harmonic, the second term on the right-hand side is identically zero and we have

$$\nabla \cdot (\phi \nabla \phi) = |\nabla \phi|^2. \quad (2.8)$$

Let us now integrate $\nabla \cdot (\phi \nabla \phi)$ over the volume V and apply the Gauss divergence theorem, namely

$$\int_V \nabla \cdot (\phi \nabla \phi) dV = \int_{\Sigma} (\phi \nabla \phi) \cdot \mathbf{n} d\Sigma.$$

The right-hand side of this equation is zero because ϕ is zero at the points of Σ , by assumption. Hence, using Eq. (2.8), we have

$$\int_V |\nabla \phi|^2 dV = 0.$$

The integrand on the left-hand side cannot be negative. Hence, the equation implies that

$$\nabla \phi = 0$$

in the entire V . This means that ϕ is uniform in the entire volume, the surface included. Being zero on the surface, ϕ is zero in the entire volume.

In conclusion, if the electrostatic potentials of the hollow conductor and of all the internal conductors are zero, or equal (considering that the potential is defined modulo an additive constant), then the entire internal region is equipotential and the field is zero. The case of the empty cavity is a particular case. We have retrieved the result discussed at the beginning of the section.

We are now ready to show that the solution of the Laplace equation with given boundary conditions is unique. Let us assume knowing a solution ϕ of the Laplace equation and that another solution, say ψ , exists with the same boundary conditions. Namely, the following conditions are satisfied

$$\begin{aligned}\phi(x, y, z) &= \psi(x, y, z) = \phi_1 && \text{on the points of } \Sigma_1 \\ \phi(x, y, z) &= \psi(x, y, z) = \phi_2 && \text{on the points of } \Sigma_2 \\ &\dots\dots\dots \\ \phi(x, y, z) &= \psi(x, y, z) = \phi_0 && \text{on the points of } \Sigma_0.\end{aligned}$$

Now, the Laplace equation being linear, $\phi - \psi$ is a solution as well, with boundary conditions

$$\phi(x, y, z) - \psi(x, y, z) = 0 \quad \text{on the points of any } \Sigma_i.$$

Hence, for the just demonstrated theorem, $\phi - \psi$ is identically zero in the entire V . ϕ and ψ are the same function.

The fact that the solution is uniquely defined by fixing the electrostatic potentials of the conductors has important practical consequences. One of these is the already-mentioned electrostatic shielding. An electrostatic shield, which is a hollow conductor, divides the space, from what concerns the electrostatic phenomena, into two completely separate and independent regions: the internal and the external.

Finally, we notice that the uniqueness theorem also holds when the surface of the enclosing conductor goes to infinity. More precisely, the theorem is also valid when one of the boundary conditions is at infinity.

2.5 Equilibrium in an Electrostatic Field

One might wonder whether it is possible to find any static arrangement of electric charges such as to produce a *stable* equilibrium position. If it was possible, we could put a charge in that position and have it remain there at rest. As we shall now show, the answer is negative. No stable equilibrium position exists in an electrostatic field. Note that this is a consequence of the inverse square dependence of the electrostatic force. As such, the conclusion is also valid for the gravitational force. There is no stable equilibrium position in the gravitational field either. We cannot put a spacecraft at some point and have it standing there in equilibrium.

A stable equilibrium position for a positive or negative charge should be in a minimum or a maximum, respectively, of the potential. We shall now show that such points do not exist for a harmonic function. Saddle points do exist, such as, for example, the point halfway between two equal point charges of the same sign that we noticed with reference to Fig. 1.11b. Indeed, this is an equilibrium position, but

the equilibrium is not stable. The potential there has a minimum moving in one direction, and a maximum moving in a direction 90° from it.

Let us assume, in a *reductio ad absurdum* argument, the harmonic function ϕ to have a maximum at the point A. Let \mathbf{r} be the position vector drawn from A. If the assumption is true, we can always find a sphere centered at A such that, at all the points on its surface, which we call S , $\partial\phi/\partial r < 0$. Thus, it will also be

$$\int_S \frac{\partial\phi}{\partial r} dS < 0.$$

Let \mathbf{n} be the unit vector normal to the sphere pointing outside. This is also the direction of \mathbf{r} on the surface, and hence, we have $\partial\phi/\partial r = \nabla\phi \cdot \mathbf{n}$. Applying the divergence theorem, we obtain

$$\int_S \frac{\partial\phi}{\partial r} dS = \int_S \nabla\phi \cdot \mathbf{n} dS = \int_V \nabla \cdot \nabla\phi dV = \int_V \nabla^2\phi dV = 0$$

where, in the last step, we took into account that ϕ is harmonic in V . Hence, the opening statement must be false.

2.6 Electrostatic Capacitance

A conductor, think of a metal to be concrete, is considered isolated if it is far enough from any other conductor and any other charged body (even if it is not a conductor). This condition is very rarely met in practice, but is easy to analyze, and we shall do that as a starting point.

Let Σ be the surface of the conductor and \mathbf{n} the unit vector of the outside normal, as shown in Fig. 2.10. Let us put the charge Q on the conductor and let ϕ_0 be the potential it takes and $\phi(x, y, z)$ the potential in the space outside the conductor. The function $\phi(x, y, z)$ is harmonic with the boundary conditions

$$\phi(x, y, z) = \phi_0 \quad \text{on } \Sigma; \quad \phi(x, y, z) = 0 \quad \text{at infinity.} \quad (2.9)$$

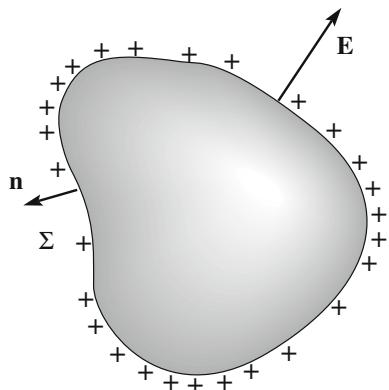
Once $\phi(x, y, z)$ is known, the field on the surface is known too, being given by

$$\mathbf{E} = -\frac{\partial\phi}{\partial n} \mathbf{n}. \quad (2.10)$$

We also know the charge density, namely

$$\sigma = \epsilon_0 E \quad (2.11)$$

Fig. 2.10 A charged isolated conductor



and the total charge of the conductor

$$Q = \int_{\Sigma} \sigma d\Sigma \quad (2.12)$$

The Laplace equation being a linear equation, if ϕ is the solution to a problem with certain boundary conditions, namely certain values of ϕ on the surfaces of the boundary (the surface of our conductor and infinity), a solution is also ϕ multiplied by any constant λ , with boundary values λ times the previous ones. In the new solution, the field will be λ times larger in every point and, and so will the charge density on the surface of the conductor and, finally, so will its charge. Inverting the argument, if we change the charge of the isolated conductor by a factor, its potential will change by the same factor, namely

$$Q = C\phi_0 \quad (2.13)$$

The proportionality constant C is the *electrostatic capacitance* or simply *capacitance* of the conductor. The name comes from the era in which electric charge was thought to be a sort of fluid and the fact that the higher the capacitance, the higher the “capacity” of the conductor to store charge at a given voltage. To be precise, the higher the capacitance, the lower the potential reached by the conductor for a given charge. The measurement unit for capacitance is the *farad* (F), after Michael Faraday (UK, 1791–1867), who made enormous contributions to all sectors of electromagnetism. The physical dimensions are coulomb per volt, namely C/V. An isolated conductor has a capacitance of one farad if, when charged with one coulomb, it reaches the potential of one volt.

One farad is a very large capacitance. To see that, let us consider a spherical conductor and let us calculate the value of its radius, say R , to have the capacitance of one farad.

Let Q be the charge on the sphere. The potential at a point immediately outside the surface, and, by continuity, on the surface as well, is the potential of a point charge at the center, namely

$$\phi_0(R) = \frac{1}{4\pi\epsilon_0} \frac{Q}{R}.$$

The capacitance is then

$$C = 4\pi\epsilon_0 R \quad (2.14)$$

To have $C = 1$ F, the radius of the sphere must be $R = 9 \times 10^9$ m, namely nine million kilometers. In practice, the submultiples are used (μF , nF , pF , etc.).

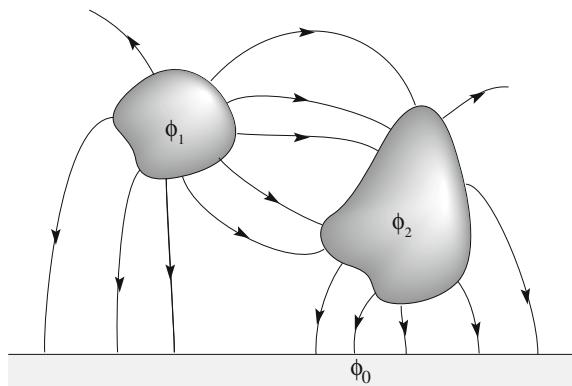
QUESTION Q 2.2. Find the capacitance of earth (radius-6400 km). How much does its potential vary if its charge increases by 1 C. \square

The concept of capacitance for an isolated conductor has a very limited practical utility, because other conductors are always present, like the walls and the floor of the room and the bodies of the people around. Under these conditions, electrostatic induction takes place. The manner in which charges distribute along the surfaces of the conductors and the potentials the conductors assume influence one another.

The simplest case consists of two conductors near one another. Even in this case, however, the presence of a third conductor, floor and walls, which we collectively call ground, cannot be ignored, as shown in Fig. 2.11. Let us assume the two conductors initially not to have a charge and to charge up them by transferring the charge Q from the conductor we shall call 2 to conductor 1. The charge of 1 and 2 will be Q and $-Q$, respectively. Let ϕ_1 and ϕ_2 be the two potentials relative to ground, whose potential we define as $\phi_0 = 0$.

Under these conditions, we have electrostatic induction between the two conductors and between each of them and ground. As opposed to the isolated conductor, there is no proportionality between the potentials of either of the two conductors, or their potential difference, and their charge. Electrostatic induction is

Fig. 2.11 Two conductors near one another



visualized by the behavior of the field lines, as in Fig. 2.11. In particular, not all the field lines leaving conductor 1 reach conductor 2. We say that the induction between them is not complete. The induction between two conductors is complete when all the field lines that exit from one enter the other. The necessary condition for that is that the charges of the two conductors are equal and opposite. The condition is not sufficient, as we have just discussed.

Looking at Fig. 2.11, it is clear that we might have complete induction by moving the pair of conductors very far from the ground. This is, however, impossible in practice, as it was for the isolated conductor. With two conductors, however, we have the possibility, which we did not have with just one. The solution is to have one of the conductors, say 1, be hollow and to lodge the second one in its cavity, as in Fig. 2.12. In this configuration, the Gauss law requires that all the field lines leaving conductor 2 terminate on the internal surface of conductor 1, because the charges on the two surfaces must be equal and opposite. The induction between 1 and 2 is complete. Note that this conclusion is independent of the charge on the external surface of conductor 1 and of the presence of conductors in the external surroundings. All of that is irrelevant for the field and the charges inside the cavity.

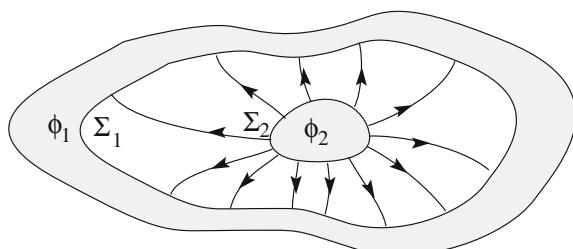
A *capacitor* is defined as a system of two conductors between which the electrostatic induction is complete. The two conductors are called the plates of the capacitor. While the word *condenser* is often used as being synonymous with capacitor, we shall use the latter terminology.

Let us now “charge up the capacitor”, meaning that we move a certain charge, say Q , from conductor 2 to conductor 1. Their charges will be $Q_1 = Q$ and $Q_2 = -Q$ and their potentials, say, ϕ_1 and ϕ_2 . We now show that the potential difference $\phi_2 - \phi_1$ is proportional to Q .

The argument is similar to that which we made for an isolated conductor. Let Σ_1 and Σ_2 be the surfaces of the two conductors. The potential $\phi(x, y, z)$ in the space between them is given by the solution of the Laplace equation with the boundary conditions $\phi(x, y, z) = \phi_1$ on Σ_1 and $\phi(x, y, z) = \phi_2$ on Σ_2 . Once $\phi(x, y, z)$ is known, we get the electric field, which is its gradient and the charge densities on the surfaces, say σ_1 and σ_2 . The charges on the conductors are then

$$Q_1 = Q = \int_{\Sigma_1} \sigma_1 d\Sigma; \quad Q_2 = -Q = \int_{\Sigma_2} \sigma_2 d\Sigma \quad (2.15)$$

Fig. 2.12 A capacitor, namely two conductors with complete induction



Again, if ϕ is the solution to a problem with the above boundary conditions, $\lambda\phi$ is the solution with boundary conditions $\lambda\phi_1$ on Σ_1 and $\lambda\phi_2$ on Σ_2 . The field near the surfaces is λ times larger and so are the surface charge densities, and so, finally, are the charges on the plates, and we can write

$$Q = C(\phi_2 - \phi_1) \quad (2.16)$$

where the constant C is the capacitance of the capacitor (or capacity of the condenser). The measurement unit of the capacitance of a capacitor is obviously the farad.

Capacitors are important elements of electric and electronic circuits. In practice, they are built joining two conductors separated by an insulating sheet (which also has the effect of increasing the capacitance, as we shall see in Chap. 4). The geometric dimensions of the surfaces of the two conductors facing one another are very large compared to their distance, in order to minimize the effects of the lack of complete closure at the borders. Indeed, in these regions, a few field lines might “escape” and end up on another conductor nearby, making the induction incomplete. In practice, however, capacitors can be produced in which these effects are negligible.

Figure 2.13 shows the capacitor of the simplest geometry, namely the parallel-plate capacitor. It is made of two equal metallic plane plates of surface S separated by a small gap of height h , which is much smaller than the diameter of S . We shall make the approximate assumptions that the field is uniform between the plates, with magnitude σ/ϵ_0 , and zero outside, as in Fig. 2.14a.

The potential difference V is the line integral of the field from one plate to the other, which is simply Eh . If Q is the charge (on the positive plate), the charge density is $\sigma = Q/S$. Hence, we have

$$V = \phi_2 - \phi_1 = \frac{\sigma}{\epsilon_0} h = \frac{Q}{S\epsilon_0} h. \quad (2.17)$$

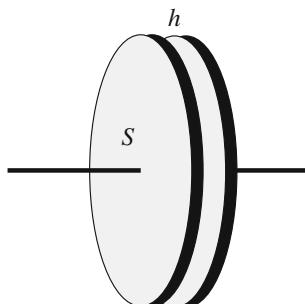


Fig. 2.13 Parallel-plate capacitor

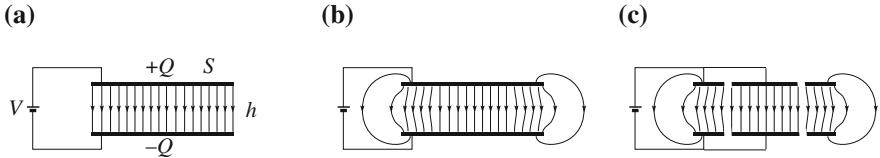


Fig. 2.14 The field of a parallel plate capacitor. **a** Ideal case, **b** as in real life, **c** with the Thomson guard-ring

In conclusion, the capacity of the parallel plate capacitor is

$$C = \frac{S\epsilon_0}{h}. \quad (2.18)$$

We see that the capacitance is greater the larger the area and the smaller the distance between the plates. To get an idea of the orders of magnitude, let us consider a parallel-plate capacitor of 1 F capacity with the distance between the plates being $h = 0.1$ mm. We immediately see that the surface needed for that is $S = 11 \times 10^6 \text{ m}^2$, namely a square with more than 3 km sides.

Equation (2.18) tells us that, as we anticipated in Sect. 1.2, we can measure the vacuum permittivity in farads per meter. In round figures, its value is, as in Eq. (1.9),

$$\epsilon_0 = 8.8 \text{ pF/m}. \quad (2.19)$$

To fix the orders of magnitude, it is good to remember that 1pF is about the capacitance of a parallel-plate capacitor of 1 cm² area and 1 mm plate separation.

Let us inquire into the validity of our assumptions. Contrary to them, the actual field does not terminate abruptly at the rim of the plates, but rather extends into the region surrounding the capacitor, as shown in Fig. 2.14b. In addition, the actual field is not uniform between the plates near the rim. Therefore, the solution we have found is not completely correct. We say that there are “fringing effects” at the edge of the capacitor. The smaller the separation h of the plates relative to their area S , the smaller the fringe effects. However, our solution can be approximated very well by the simple modification devised by William Thomson (UK, 1824–1907). We divide both plates into a central part, where the field is uniform, and in an external “guard ring” separated by a very narrow gap. The ring is in the same plane and has the same potential as the nearby plate, as shown in Fig. 2.14c. Our capacitor is now the central part of the system, from which the edge effects are removed. Equation (2.18) holds with a very good approximation.

2.7 Calculating Capacitances

In this section, we shall calculate the capacitances of capacitors of two symmetric geometries, spherical and cylindrical.

Spherical capacitor.

Figure 2.15 shows a cross-section through the center of a spherical capacitor, which consists of two concentric spherical conducting shells, one inside the other. A small hole is made in the outer sphere to allow an electric connection with the inner one going through. Let R_1 and R_2 be the radii of the inner and outer conductors, respectively, and Q the charge of the capacitor.

The field between the two surfaces is the field of a point charge Q in the center, namely

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0} \frac{Q}{r^2} \mathbf{u}_r$$

The potential difference is obtained by integrating the field on a line between the plates, which we chose to be along a radius. We then have

$$\phi_2 - \phi_1 = - \int_{R_1}^{R_2} \mathbf{E} \cdot d\mathbf{r} = - \frac{Q}{4\pi\epsilon_0} \int_{R_1}^{R_2} \frac{dr}{r^2} = \frac{Q}{4\pi\epsilon_0} \left(\frac{1}{R_2} - \frac{1}{R_1} \right),$$

which gives the capacitance

$$C = 4\pi\epsilon_0 \frac{R_1 R_2}{R_2 - R_1}. \quad (2.20)$$

We observe that if the plates are very near, namely if $R_1 \approx R_2$, and we call $h = R_2 - R_1$ the distance between the conductors, then we have $C = \epsilon_0 4\pi R^2 / h = \epsilon_0 S / h$, which is the capacitance of the parallel plate capacitor.

Fig. 2.15 Equatorial cross-section of a spherical capacitor

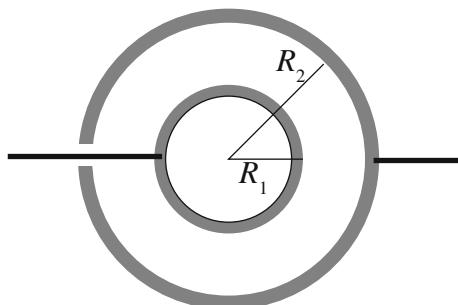
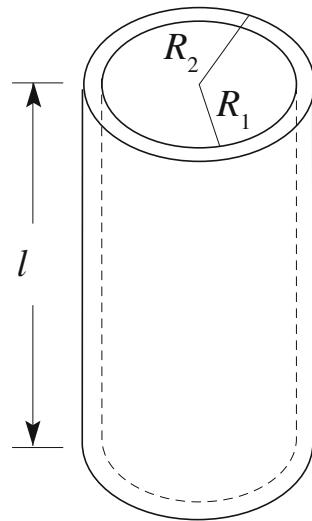


Fig. 2.16 Geometry of a cylindrical capacitor



In order to compare this with the spherical isolated conductor, let us calculate the radius of a spherical capacitor of 1 F with 0.1 mm spacing between the plates. We find $R = 300$ m, which is quite big, but not really enormous.

Cylindrical capacitor.

The cylindrical capacitor is made of two coaxial conducting shells, one inside the other. Let R_1 and R_2 be the radii of the inner and outer conductors, respectively, l their height and Q the charge of the capacitor. Figure 2.16 shows the geometry.

The field between the electrodes is the field of a linear charge distribution on the axis with linear density $\lambda = Q/l$. Its magnitude at the distance r' from the axis is

$$\mathbf{E} = \frac{\lambda}{2\pi\epsilon_0 r'} \mathbf{u}_{r'}. \quad (2.21)$$

By integration, we obtain the potential difference

$$\phi_2 - \phi_1 = - \int_{R_1}^{R_2} \mathbf{E} \cdot d\mathbf{r}' = - \frac{\lambda}{2\pi\epsilon_0} \ln \frac{R_2}{R_1} = - \frac{Q}{2\pi\epsilon_0 l} \ln \frac{R_2}{R_1}.$$

The capacitance is then

$$C = 2\pi\epsilon_0 l / \ln(R_2/R_1). \quad (2.22)$$

We leave as an exercise to show that if $R_1 \approx R_2$, this expression reduces to the one valid for the parallel plate capacitor with $h = R_2 - R_1$. Use the approximation $\ln(R_2/R_1) = \ln[(R_1 + h)/R_1] = \ln(1 + h/R_1) \approx h/R_1$.

In practice, a cylindrical capacitor is build by overlapping two rectangular flexible metal strips separated by an insulating layer. To have a handy device, one side of the rectangle is short, 1–2 cm, the other very long. The sandwich is then wrapped in a helix to form a cylinder.

In practice, as in the just-considered example, capacitors have an insulator rather than a vacuum between the conductors. Electric insulators are also called dielectrics. We shall study their properties in Chap. 4. We shall see there that the expressions we have found for the capacitances need to be modified, simply changing the vacuum permittivity ϵ_0 into a constant ϵ characteristic of the medium, called the permittivity of the material. A connected term is the relative permittivity, also called the dielectric constant that is the ratio $\kappa = \epsilon/\epsilon_0$. The dielectric constant is, in any case, larger than one, having values ranging, for different media, from a few units to hundreds of thousands.

Another important feature of the dielectric used to separate and insulate the plates is its dielectric strength. This is the maximum field that the material can withstand without breaking down. Breakdown results in the formation of an electrically conductive path and a discharge through the material. For a solid material, a breakdown generally destroys its insulating capability. Good insulators have dielectric strengths up to tens of MV/m.

2.8 Combining Capacitors

As we already mentioned, capacitors are commonly used in electronic circuits, sometimes in quite complicated combinations. It is thus useful to have a set of rules for finding the *equivalent capacitance* of the different combinations of capacitors. It turns out that we can always find the equivalent capacitance by repeated application of two simple rules. These rules are for the two basic connection types: in series and in parallel.

Figure 2.17 shows n capacitors connected in series. The arrangement forms a line in which the positive plate of a capacitor is connected to the negative plate of the next one. The two plates become a unique conductor. Now, let us charge the line by taking a charge Q from the last plate on the right (which will then have the charge $-Q$) and putting it on the first plate on the left (which will then have the charge $+Q$). The charge on the other plate of the first capacitor is, by induction, $-Q$. No net charge has gone on the conductor made of the right plate of the first capacitor and the left plate of the second. Its net charge is zero. Consequently, the left plate of the second capacitor has the charge $+Q$. The process repeats itself to the end of the line.

The potential difference on the i -th capacitor is $\Delta\phi_i = Q/C_i$. The total potential difference between the extremes of the line is then

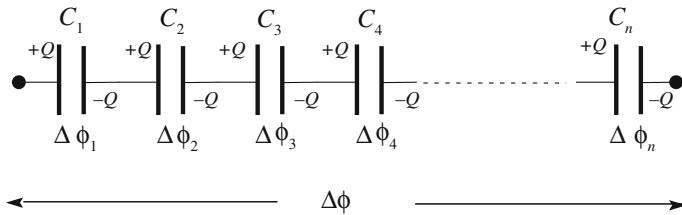


Fig. 2.17 Capacitors connected in series

$$\Delta\phi = \sum_{i=1}^N \Delta\phi_i = Q \sum_{i=1}^N \frac{1}{C_i},$$

which we see to be proportional to the charge. This means that a series of capacitors behaves like a single capacitor with the equivalence capacitance

$$C = 1 / \sum_{i=1}^N \frac{1}{C_i}. \quad (2.23)$$

Figure 2.18 shows n capacitors connected in parallel. In this arrangement, all the positive plates are connected together, thus forming a unique conductor, and the negative plates are similarly connected.

Consequently, the potential difference is the same for all capacitors. Let it be $\Delta\phi$ and let Q be the total charge on the connected plates. This charge distributes on the capacitors depending on their capacitance. Indeed, the charge on the i th capacitor is $Q_i = \Delta\phi \times C_i$. Adding them up, we have

$$Q = \Delta\phi \sum_{i=1}^N C_i.$$

Even now, the charge is proportional to the potential difference and we can state that a system of capacitors connected in parallel is equivalent to a single capacitor of equivalent capacitance

$$C = \sum_{i=1}^N C_i. \quad (2.24)$$

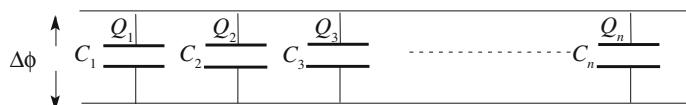


Fig. 2.18 Capacitors connected in parallel

To deal with complicated arrangements of capacitors, one starts by considering subsets in which the capacitors are connected in series or in parallel. To each subset, one substitutes a single equivalent capacitor using the above rules. The arrangement is now simpler and we can apply the same procedure over and over again, since we are left with a single equivalent capacitor.

2.9 Electrostatic Induction Coefficients

In Sect. 2.6, we considered a system consisting of a hollow conductor containing in its cavity a second conductor. We have seen that the electrostatic induction between them is complete and introduced the concept of capacitance. In this section, we shall consider a more general situation in which the hollow conductor contains any number of conductors in its cavity.

To be concrete, we shall consider a system of three conductors in the cavity, without affecting the generality of the argument. Let Q_1 , Q_2 and Q_3 be the charges of the three internal conductors and ϕ_1 , ϕ_2 and ϕ_3 their potentials. Let Q_0 be the charge on the internal surface of the hollow conductor (the charge on its external surface is irrelevant for the field in the cavity) and ϕ_0 its potential. As we know, $Q_0 = -Q_1 - Q_2 - Q_3$. Let us search for the relation between charges and potentials. We shall use the superposition principle and the uniqueness of the solution of the Laplace equation.

We start by considering, one after the other, the three particular arrangements shown in Fig. 2.19. One case is when $\phi_2 = \phi_3 = \phi_0$ and ϕ_1 is arbitrary. We can consider connecting conductors 1 and 2 to the hollow one with two conductive wires. We call this configuration state 1 of the system. Let Q_1^1 , Q_2^1 and Q_3^1 be the charges of the three internal conductors (the superscript indicates the state we are considering). One can easily see that the three charges are proportional to $\phi_1 - \phi_0$. The argument is the same one we have already used a few times. If, for example, $\phi_1 - \phi_0$ were to double, the field would double too, and the charges Q_1^1 , Q_2^1 and Q_3^1 as well. We can then write

$$Q_1^1 = C_{11}(\phi_1 - \phi_0); \quad Q_2^1 = C_{21}(\phi_1 - \phi_0); \quad Q_3^1 = C_{31}(\phi_1 - \phi_0).$$

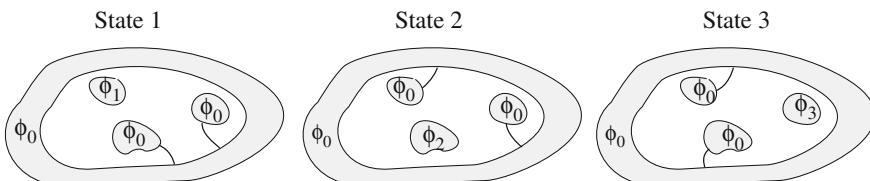


Fig. 2.19 Three states of the system of three conductors in a cavity

Let us now consider State 2, in which $\phi_1 = \phi_3 = \phi_0$ and ϕ_2 is arbitrary. The charges, which we call Q_1^2 , Q_2^2 and Q_3^2 , are proportional to $\phi_2 - \phi_0$, namely

$$Q_1^2 = C_{12}(\phi_2 - \phi_0); Q_2^2 = C_{22}(\phi_2 - \phi_0); Q_3^2 = C_{32}(\phi_2 - \phi_0).$$

Finally, State 3 is when $\phi_1 = \phi_2 = \phi_0$ and ϕ_3 is arbitrary. The charges, Q_1^3 , Q_2^3 and Q_3^3 , are proportional to $\phi_3 - \phi_0$, namely

$$Q_1^3 = C_{13}(\phi_3 - \phi_0); Q_2^3 = C_{23}(\phi_3 - \phi_0); Q_3^3 = C_{33}(\phi_3 - \phi_0).$$

The proportion coefficients C_{ij} are called electrostatic induction coefficients and have the physical dimensions of a capacitance. If there is only one conductor, say i , we are back to the case in Sect. 2.6. The system is a capacitor and C_{ii} is its capacitance.

The general case in which the potentials are arbitrary is immediately obtained considering the superposition of States 1, 2 and 3. The Laplace equation being linear, the potential difference between any internal point and the hollow conductor is the sum of the corresponding potential differences in the three cases, the field is the sum of the fields and the charge on each surface the sum of the charges. Adding up the three equations and calling $Q_i = Q_i^1 + Q_i^2 + Q_i^3$, we have

$$\begin{aligned} Q_1 &= C_{11}(\phi_1 - \phi_0) + C_{12}(\phi_2 - \phi_0) + C_{13}(\phi_3 - \phi_0) \\ Q_2 &= C_{21}(\phi_1 - \phi_0) + C_{22}(\phi_2 - \phi_0) + C_{23}(\phi_3 - \phi_0). \\ Q_3 &= C_{31}(\phi_1 - \phi_0) + C_{32}(\phi_2 - \phi_0) + C_{33}(\phi_3 - \phi_0) \end{aligned} \quad (2.25)$$

In conclusion, the charges of each conductor depend linearly on the potential differences between them and the external conductor. If we fix the arbitrary constant of the potentials, taking the potential of the hollow conductor as zero, then the charge of each internal conductor is proportional to its potential.

2.10 Electrostatic Shield

Let us again consider a system of n conductors enclosed in the cavity of an external conductor, as shown in Fig. 2.20.

We already know that, once the potentials $\phi_0, \phi_1, \phi_2 \dots \phi_n$ are fixed, the field in the cavity and the charges on the conductors are defined. As we shall now see, this implies that regardless of any change that can be made in the field and in the charges outside the external conductor, no observable change can happen inside the cavity (under static conditions). We might move the charged external bodies, such as the one in the figure, change their charges, or even put the charge from the outside on the external surface of the external conductor. In the latter case, its

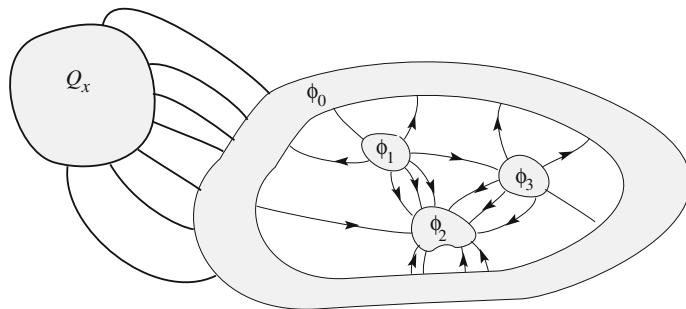


Fig. 2.20 Electrostatic shielding

potential will vary, but the potential differences with the internal conductors will remain exactly the same. We shall now prove the latter sentence.

We start by observing that whatever we can do outside the charge on the surface of the cavity cannot change because it is equal to the opposite of the charges of the internal conductors.

Let us next consider the simplest situation in which there is only one internal conductor. This is the case already studied in Sect. 2.6. The system is a capacitor and the potential difference between internal and external conductors, say $\phi_1 - \phi_0$, is proportional to the charge of the internal one. No external action, in statics, can change $\phi_1 - \phi_0$. We can change from outside the charge of the external conductor and its potential as a consequence, but the potential of the internal conductor will change by the same amount.

We can extend the same argument to the case of a number of conductors in the cavity. Indeed, as we have shown in Sect. 2.9, their charges, which cannot vary for an external action, are linked by linear relations to the potential differences with the external conductor. Hence, any external electrostatic action can only change the potentials of all the conductors, $\phi_0, \phi_1, \phi_2 \dots \phi_n$, by the same quantity.

In the volume of the cavity, the potential $\phi(\mathbf{r})$ is the solution of the Laplace equation with the boundary conditions given by the potentials $\phi_0, \phi_1, \phi_2 \dots \phi_n$ of the conductors, which form the surface of that volume. As we just saw, a change in the external conditions can change the potential, but all by the same quantity, say $\Delta\phi$. On the other hand, $\phi(\mathbf{r}) + \Delta\phi$ is also a solution of the Laplace equation. More precisely, this solution satisfies the new boundary conditions and consequently, the solution being unique, it is *the* solution. In conclusion, a change in the external electrostatic conditions can change the potential at all the points of the cavity by the same additive amount. But any observable effect is due to the field, not to the potential, and the field does not vary if the potential varies at every point by the same additive quantity. This is the action of electrostatic shielding already mentioned.

Notice that the shielding action works in both direction. Namely, whatever we may do inside (electrostatically), like moving charges from one conductor to the other, surface of the cavity included, no effect can be observed outside. In electrostatics, internal and external spaces are completely separate and independent.

We can also express our conclusions by stating that, under static conditions, an observer inside the cavity of a conductor cannot determine with any measurement the potential of the conductor that encloses him/her, which we can call the potential of his/her laboratory. In other words, the laws of the electrostatic phenomena are *invariant* under a change of the potential of the laboratory, namely they have the *same form* for whatever that potential is. This is the particular form met in electrostatics of a general property of electromagnetism called *gauge invariance*.

Electrostatic shielding has several practical applications. Delicate experiments often employ very sensitive equipment that is affected by the presence of conductors in the surroundings, especially when measuring very small charges or potential differences. To protect them, one encloses the instruments in a metallic, grounded structure, often in the shape of a cage, called a Faraday cage, having been invented in 1836 by Michael Faraday (UK, 1791–1867). The walls of the structure can be made of a continuous metal foil but also of a wire mesh. The latter solution works because it can be shown that the influence of the openings between the wires of the mesh only extends over distances comparable with their diameters. Faraday cages are capable of providing a partial shield for fields variable in time as well.

Note that the properties discussed in this chapter ultimately stem from the inverse square dependence of the electrostatic force. If the field of a point charge varied with a power of distance different from -2 , even of a very small amount, the Gauss law would not be valid and, in particular, the shielding action would not be complete. This feature provides an extremely sensitive way to measure the difference from -2 of the r exponent.

Suppose that the dependence on the distance of the force is $1/r^{-2+\varepsilon}$. We build a hollow conductor, for example, a spherical shell. We charge it from outside, raising its potential as much as we can, to obtain the maximum effect, if any. We measure the charge on the internal surface with a sensitive electrometer. If we do not find any charge, as is the case, we can say that if any charge is present, it must be smaller than the sensitivity of our instrument (we can never say that it is exactly zero). In parallel, we calculate how much charge we would expect to find as a function of ε . The upper limit found on the charge will then translate into an upper limit of ε . Notice that, while, in principle, any shape of hollow conductor will do, having a symmetric shape, namely a sphere, makes the calculation possible in practice. Notice also that we shall never be able to make a *perfect* sphere, and that, consequently, we need to take that into account in evaluating the experimental uncertainties.

The invention of the method is credited to Joseph Priestley (UK, 1773–1804) in 1767. In that year, which, it should be noted, was 18 years before the Coulomb experiment, the idea already existed that the electrostatic force might have inverse square law dependence in analogy to the gravitational force. Priestley knew the Newton theorem showing that the gravitational field inside a spherical shell is zero (see Vol. 1, Sect. 4.6). He measured the charge inside a spherical shell, finding none, within a limited sensitivity. This was the first historical hint at the inverse square law, which was later established by Coulomb with a direct measurement (Sect. 1.2).

The first precise measurement with Priestley's method was done by Henry Cavendish (UK, 1731–1810), who, in 1773, established the limit $|\epsilon| \leq 0.02$. One century later, in 1873, James Clerk Maxwell (UK, 1831–1879) improved the limit to $|\epsilon| \leq 5 \times 10^{-5}$. The limit constantly improved over time until it reached the present value of $|\epsilon| \leq 6 \times 10^{-17}$. Here, we mention that, in quantum mechanics, the electromagnetic interaction is described as being due to the exchange of photons, which are the quanta of the electromagnetic field. The mass of the photon is rigorously zero if the exponent is exactly -2 . Consequently, the upper limit on $|\epsilon|$ provides an upper limit to the photon mass. Taking the electron mass m_e for comparison, $|\epsilon| < 6 \times 10^{-17}$ corresponds to the photon mass m_γ , being such that $m_\gamma/m_e < 10^{-21}$.

2.11 The Method of Images

The solution of the Laplace equation is unique not only if the boundary conditions are on surfaces at finite distances, but also if they are at infinity. It can be shown, although we shall not do that, that this is true if both potential and field go to zero fast enough when the distance goes to infinity.

The problem in finding the solution is usually much more difficult, depending on the boundary conditions. As a matter of fact, the only general methods are numerical, using powerful computer codes. In a few particularly simple cases, however, the solution can be found with certain “tricks”. One of these is the method of images that we shall now see.

Let us start with an electric field generated by a charge arrangement that we are able to calculate. For example, Fig. 2.21 represents the field of two equal and opposite point charges. Let us now take a metal sheet and give it the form of one of the equipotential surfaces, placing it exactly on that surface and giving it the potential of the surface. The sheet divides the space into two regions, separated from one another by the electrostatic shielding of the sheet. In each of them, the field is exactly the same as before the introduction of the sheet, because the boundary conditions in each of them have not changed. We can now fill one of the two regions with a conducting medium without inducing any change in the other one. We have thus found a solution to the Laplace equation for a system consisting of a point charge near a solid conductor with a certain surface shape (the shape of the equipotential we have chosen) and a certain potential. Considering a number of cases like this, we end up with a collection of solutions to possible problems. If we should encounter one of these problems, we can pick the solution from our collection.

The simplest of such problems is shown in Fig. 2.22. The challenge is to find the field of a point charge q at the distance z_q from a grounded plane conductor. Looking at Fig. 2.21, we see that one of the equipotential surfaces, having null potential, is the middle plane between the two charges. We then place a plane conductor at zero potential (grounded) on that plane. The field in the semispace on

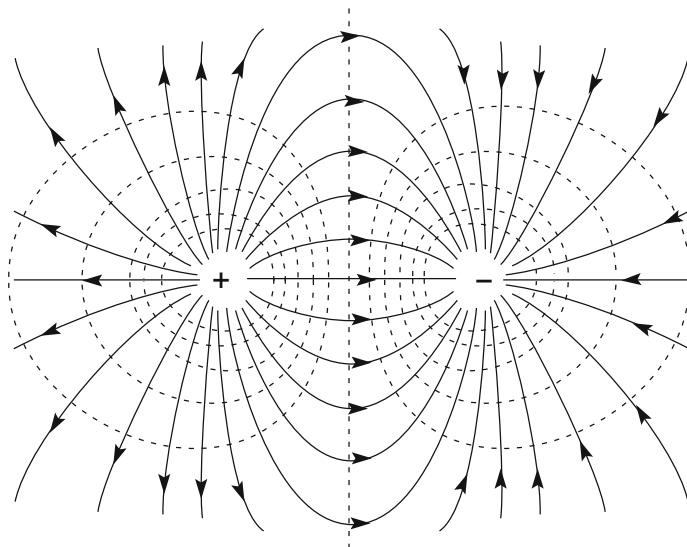
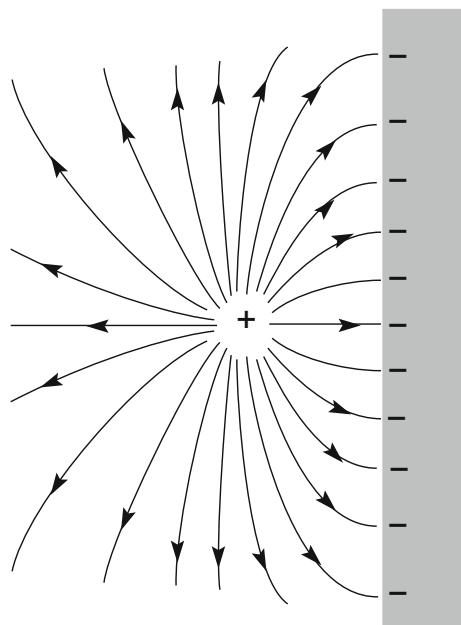


Fig. 2.21 Electric field lines and equipotential surfaces of two equal and opposite charges

Fig. 2.22 Electric field lines for a positive charge and a conductive plane



the left of this conductor is equal to the field that there would be without the plane, but with a charge $-q$ at $-z_q$. This is the specular distance relative to the real charge q and is called an image charge of q . As a matter of fact, the real charge sources of

the field are q and the negative charges that move inside the conductor to its surface and arrange themselves to make the field inside null. As we know the field at every point on the surface, we also know the charge density σ , which is the field magnitude divided by ϵ_0 .

Let us take a reference frame with the origin O at the point of the conductor plane below the charge, the z -axis normal to the surface through q , and the x and y axes on the surface. The electric field on the surface has the direction equal and opposite to the z -axis. Let $\rho = \sqrt{x^2 + y^2}$ be the distance from O of a generic point (x, y) of the surface. The field at that point due to q and its image is

$$E_z = -\frac{1}{4\pi\epsilon_0} \frac{2z_q q}{(z_q^2 + \rho^2)^{3/2}}.$$

The charge density is then

$$\sigma = -\frac{1}{2\pi} \frac{z_q q}{(z_q^2 + \rho^2)^{3/2}}. \quad (2.26)$$

We leave as an exercise the verification that the integral of the surface density on the plane is just $-q$.

We finally observe that a charge q facing a conductor is acted upon by an attractive force due to the negative charges it induces on the surface of the conductor. In the case just discussed, the force can be evaluated as the attractive force of the image charge, namely

$$F_z = -\frac{1}{4\pi\epsilon_0} \frac{q^2}{(2z_q)^2}$$

In similar, more complicated cases, image charges are always present, but in general, their values and their distances from the surface are not equal and opposite to those of the real charge. These quantities must be evaluated case by case.

QUESTION Q 2.3. A nucleus of Fe ($Z = 26$) is at rest at 1 μm from a plane grounded conductor. Is there any force on the nucleus? If yes, what is its value? \square

Summary

In this chapter, we have learned the following principal concepts:

1. The concept of the electric conductor.
2. The properties of a conductor under static conditions; electric field and charge density inside the conductor and on its surface.

3. The force on the surface of a charged conductor.
4. The properties of hollow conductors; electrostatic shielding.
5. How to calculate the potential, and the field, in a space region between conductors at fixed potentials, when the charges that produce the field are not given.
6. That no stable equilibrium for a charge exists in an electrostatic field.
7. The concept of electrostatic capacitance.
8. Capacitances and their different arrangements in electronic circuits.
9. The method of images.

Problems

- 2.1. How does the electrostatic field vary doubling the distance from its sources if they are: (a) a point charge, (b) a linear uniform charge distribution, (c) a planar uniform charge distribution?
- 2.2. Two metal spheres of radii R_1 and R_2 at a distance much larger than the radii are connected by a conducting wire. The system is charged with the charge Q . Find the charges Q_1 and Q_2 on the two spheres.
- 2.3. The capacitance of a metallic conductor depends or does not depend on (choose): the metal, the shape, the temperature, the presence and the position of other conductors.
- 2.4. Does any charge exist inside a conductor? Which is the charge density in a conductor under static conditions?
- 2.5. You are inside a Faraday cage that is on insulating supports. You know that the cage is connected outside to a constant voltage generator of 100 kV, but you do not know if the generator is on or off. Would you touch the wall of the cage?
- 2.6. Fig. 2.23 is a cartoon showing hypothetical lines of an electric field in the presence of three conductors, A having a positive charge, B having a negative charge and C having no charge. The figure has 7 different mistakes. Find them and explain.
- 2.7. The capacitance of a cylindrical capacitance depends or does not depend on (choose): the radii of its surfaces, their heights, the metal of which it is composed, its charge, the presence of other conductors.
- 2.8. Four spherical drops of water are connected to the positive pole of a battery having the potential ϕ and then disconnected. They then merge into a single drop. What is its potential?
- 2.9. Four plane square conductors form four faces of a cube and have potentials as in Fig. 2.24. Draw the electric field lines.
- 2.10. Two plane-conducting surfaces are charged with equal surface density. How does the repulsive force between them vary with their distance?

Fig. 2.23 Conductors and field lines

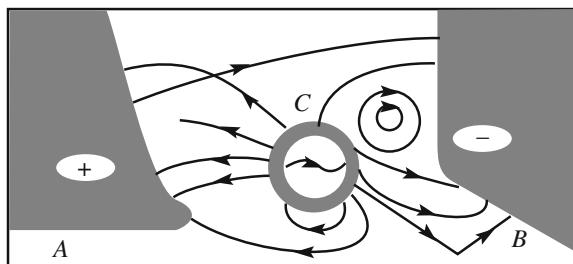


Fig. 2.24 Four plane conductors and their potentials

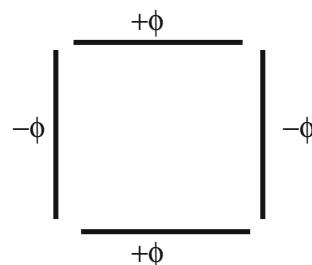
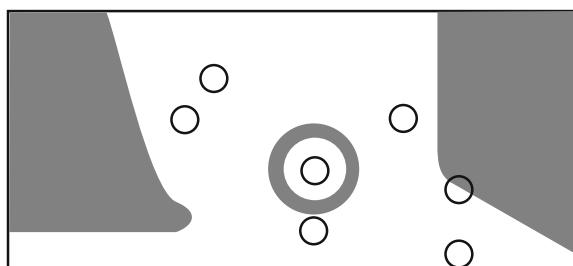


Fig. 2.25 Solution to Problem 2.7



- 2.11. The point charge $Q = 50 \text{ nC}$ is at rest at a distance of 1 mm from a grounded plane metal sheet. There are no other materials. What is the force on the charge? What is the charge density on the sheet as a function of the distance of the foot of the normal to the sheet from the charge? What is the induced charge?
- 2.12. We have two capacitances $C_1 = 10 \mu\text{F}$ and $C_2 = 35 \mu\text{F}$. We connect them in series and apply to the series the potential difference $\Delta\phi = 100 \text{ V}$. What are the potential differences at each capacitor? (Fig. 2.25).

Chapter 3

Electrostatic Energy

Abstract In this chapter, we find the expression of the potential energy of point charges, of continuous charge distributions and of charged conductors in a vacuum. We shall see how energy can be thought of as continuously distributed in space with a density proportional to the square of the field intensity. We shall find that the energy of the field of a point charge is infinite.

We have learned that the electrostatic force and the electrostatic field are conservative. In this chapter, we shall find the expression of the potential energy of relevant electrostatic systems, both of charges in a vacuum and of conductors. We shall start in Sect. 3.1 with the simplest case, which is a system of point charges in fixed and known positions in a vacuum. In Sect. 3.2, we shall deal with a continuous charge distribution. Sections. 3.3 and 3.4 deal with the energy of charged conductors.

In Sect. 3.5, we shall see how energy can be thought of as continuously distributed in space with a density proportional to the square of the field intensity. In Sect. 3.6, we address the energy of the field of point charges. We shall see that this energy is infinite, because the work required to “assemble” a finite quantity of charge at a point, namely a space region of null diameter, is infinite. Point charges do exist in nature. Such is the case, for example, with electrons, as far as we know. The problem of the energy of their field being infinite, however, has no practical consequences, because electrons are given; we have not assembled them and we cannot destroy them.

The energy of charges immersed in a dielectric will be addressed in the next chapter.

3.1 Energy of a System of Point Charges

Consider a system of three point charges q_1 , q_2 and q_3 at rest in a vacuum in the positions \mathbf{r}_1 , \mathbf{r}_2 and \mathbf{r}_3 . Let us express the energy of the system. As always, the energy of the system is the work to be done to build it, starting from the configuration in which energy is zero by definition. We define the energy to be zero in the state in which the charges do not interact, namely when their distances are infinite. Notice that the charges, each by itself, are considered as given, namely we do not include the work needed to build each point charge.

Let us start by moving the charge q_1 from infinite (very large in practice) distance to its final position \mathbf{r}_1 . There is no force yet acting on the charge and the work is zero. Let us now move the charge q_2 from infinity to \mathbf{r}_2 . In doing that, we must block q_1 , already in position, but no work is needed for that. To move q_2 , we must do a work against the force \mathbf{F}_{12} exerted by q_1 on q_2 . This is just q_2 times the potential of the field of q_1 at \mathbf{r}_2 , say $\phi_1(\mathbf{r}_2)$. Hence, the work is

$$W_2 = q_2 \phi_1(\mathbf{r}_2) = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r_{12}}$$

where $r_{12} = |\mathbf{r}_2 - \mathbf{r}_1|$.

To now bring the charge q_3 into position, we must work against the force \mathbf{F}_{13} exerted by q_1 and the force \mathbf{F}_{23} exerted by q_2 . This work is q_3 times the potential of the fields of q_1 and q_2 at \mathbf{r}_3 , namely

$$W_3 = q_3 [\phi_1(\mathbf{r}_3) + \phi_2(\mathbf{r}_3)] = \frac{1}{4\pi\epsilon_0} \left(\frac{q_1 q_3}{r_{13}} + \frac{q_2 q_3}{r_{23}} \right)$$

The total work, which is the energy of the system, is then

$$U_E = \frac{1}{4\pi\epsilon_0} \left(\frac{q_1 q_2}{r_{12}} + \frac{q_1 q_3}{r_{13}} + \frac{q_2 q_3}{r_{23}} \right). \quad (3.1)$$

Notice that this expression is symmetrical under the exchange of any pair of charges, namely it is independent of the order in which we bring the charges in, as it should be. Notice also that U_E may be both positive and negative. It is positive when, for example, all the charges have the same sign. In this case, the work to build the system is positive. The acting forces are repulsive; they tend to destroy the system, diminishing its energy. For example, all the charges in an atomic nucleus are positive (they are protons). They do not destroy the nucleus because they are balanced by the nuclear force that is attractive. It is possible in some circumstances to unlock the nuclear force. The nucleus then breaks into pieces, whose kinetic energy is equal to the initial electric potential energy. As a matter of fact, that which is usually called nuclear energy is indeed electrostatic energy.

If the potential energy is negative, at least part of the forces are attractive and tend to approach the charges and, as always, to decrease the energy. For example,

an ionic crystal, such as the common NaCl, is composed of positive (Na^+) and, alternative to them, negative ions (Cl^-). They attract each other. To destroy the system, we must give energy, for example, by heating and transforming the crystal into a gas.

We finally notice that Eq. (3.1) is a sum over the contributions of *pairs* of charges. Each term is the *interaction energy* of a pair.

Equation (3.1) is immediately extended to the general case of N point charges at rest in a vacuum as

$$U_E = \frac{1}{2} \sum_{i=1}^N \sum_{j=1; j \neq i}^N \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}. \quad (3.2)$$

Notice that in this expression: (a) we have specified $j \neq i$ in the sum, because the interactions are between different charges, not of one charge with itself; and (b) we have multiplied by $\frac{1}{2}$, because in the sum, each pair is counted twice, once as ij and once as ji .

We can write Eq. (3.2) in the equivalent form as

$$U_E = \frac{1}{2} \sum_{i=1}^N q_i \cdot \left(\sum_{j=1; j \neq i}^N \frac{q_j}{4\pi\epsilon_0 r_{ij}} \right). \quad (3.3)$$

Here, we recognize the term in parenthesis as the potential in the position \mathbf{r}_i of q_i due to all the *other* charges of the system. Calling it ϕ_i , we have

$$U_E = \frac{1}{2} \sum_{i=1}^N q_i \cdot \phi_i. \quad (3.4)$$

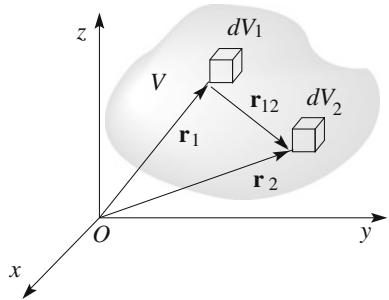
QUESTION Q 3.1. Nuclear fission can be induced in a ^{235}U by having it absorb a neutron. The resulting ^{236}U breaks in ^{92}Kr , ^{141}Ba and 3 neutrons. Give an evaluation of the energy released in the process. \square

QUESTION Q 3.2. Calculate the electrostatic interaction energy of a proton and an electron at the mean distance in a hydrogen atom, which is 80 pm. How much is the corresponding energy of one mole of hydrogen (H_2). \square

3.2 Energy of a Continuous Charge Distribution

Let us now express the energy of a continuous system of charges, at rest in a vacuum. Let V be the volume of the distribution and $\rho(\mathbf{r})$ its (constant in time) charge density. As shown in Fig. 3.1, let us consider two infinitesimal volumes dV_1 and dV_2 in the positions \mathbf{r}_1 and \mathbf{r}_2 , respectively.

Fig. 3.1 A continuous charge distribution



The interaction energy of the charges in the two volumes is clearly

$$dU_E = \frac{\rho(\mathbf{r}_1)dV_1\rho(\mathbf{r}_2)dV_2}{4\pi\epsilon_0 r_{12}}$$

The energy of the system is the double volume integral in dV_1 and dV_2 over the total volume, which should be divided by 2, for the same reason as for the discrete distribution. We then have

$$U_E = \frac{1}{2} \int_V \int_V \frac{\rho(\mathbf{r}_1)dV_1\rho(\mathbf{r}_2)dV_2}{4\pi\epsilon_0 r_{12}}. \quad (3.5)$$

Similarly to the discrete case, we can write

$$U_E = \frac{1}{2} \int_V \rho(\mathbf{r}_2) \left[\int_V \frac{\rho(\mathbf{r}_1)dV_1}{4\pi\epsilon_0 r_{12}} \right] dV_2 = \frac{1}{2} \int_V \rho(\mathbf{r}_2)\phi(\mathbf{r}_2)dV_2, \quad (3.6)$$

because

$$\phi(\mathbf{r}_2) = \int_V \frac{\rho(\mathbf{r}_1)dV_1}{4\pi\epsilon_0 r_{12}}$$

is the potential in \mathbf{r}_2 of *all* the charge distribution. This time, we did not specify “due to all the charges but those in \mathbf{r}_2 ” because the latter is infinitesimal and so is its contribution. Note also that the integral diverges at the points for which $\mathbf{r}_1 = \mathbf{r}_2$. This is no problem, as long as ρ is a regular function, because the region $\mathbf{r}_1 = \mathbf{r}_2$ is a measure zero set. However, if there are point charges, the function ρ diverges in their position. Consequently, the expressions of this section do not hold in the presence of point charges.

In conclusion, we have found that the energy of a continuous charge distribution with charge density $\rho(\mathbf{r})$ is given by

$$U_E = \frac{1}{2} \int_V \rho(\mathbf{r})\phi(\mathbf{r})dV \quad (3.7)$$

where $\phi(\mathbf{r})$ is the potential at the point \mathbf{r} generated by the charge distribution itself and V is the volume of the distribution.

3.3 Energy of a System of Conductors

Consider a system of n conductors with charges Q_1, Q_2 , etc., as shown in Fig. 3.2. We define the state of zero energy as the state in which the conductors are uncharged in their positions. Let us calculate the work to be done on the system to charge the conductors.

Let ϕ_i be the final potential and $\sigma_i(\mathbf{r})$ the surface charge density of the i -th conductor. The charges Q_i are obviously

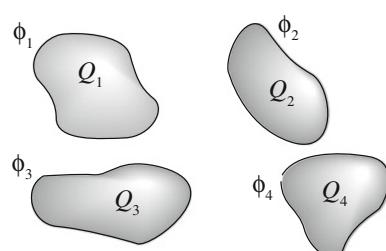
$$Q_i = \int_{S_i} \sigma_i(\mathbf{r})dS. \quad (3.8)$$

Let us express the energy, taking into account that charges are only on the surfaces S_i , as

$$U_E = \frac{1}{2} \int_V \rho(\mathbf{r})\phi(\mathbf{r})dV = \frac{1}{2} \sum_i \phi_i \int_{S_i} \sigma_i(\mathbf{r})dS,$$

where, on the right-hand side, we have taken ϕ_i out of the integral, because it does not depend on the point of the surface S_i . Considering that the integral on the right-hand side are the charges of the conductors, we finally have

Fig. 3.2 Four conductors, their charges and potentials



$$U_E = \frac{1}{2} \sum_{i=1}^N Q_i \cdot \phi_i. \quad (3.9)$$

This expression is formally equal to Eq. (3.4) holding for point charges, but it has a very different meaning. In the case of point charges, ϕ_i is the potential of all the *other* charges at the point where the i -th charge is located, while in the case of conductors, ϕ_i is the potential due to *all* the charges of the system, those of the i -th conductor included. As a matter of fact, we can reach the same result differently. We start with the conductors that are very far from one another and charge each of them to Q_i . We then take the conductors in their final positions. Both operations require work on the system. In the case of the point charges, only the second operation is needed. In the case of conductors, the work for charging each of them is additionally necessary.

3.4 Energy Stored in a Capacitor

We now calculate the energy stored in a capacitor with a given charge. The system being made of two charged conductors, we can start from the result of the last section. Let Q_1 and Q_2 be the charges, ϕ_1 and ϕ_2 the potentials of the two plates and C the capacitance. The charges are linked to the potential difference as $Q_1 = C(\phi_1 - \phi_2)$ and $Q_2 = -Q_1 = -C(\phi_1 - \phi_2)$.

Equation (3.9) then gives us $U_E = (\phi_1 Q_1 + \phi_2 Q_2)/2 = C(\phi_1 - \phi_2)^2/2$. Calling $V = \phi_1 - \phi_2$, we have

$$U_E = \frac{1}{2} CV^2 = \frac{1}{2} \frac{Q^2}{C}. \quad (3.10)$$

We can obtain this expression differently, namely by computing the electric work needed to charge the capacitor. Initially, the charges on each plate are zero. We proceed by moving charges from one plate to the other. Let q and $-q$ be the charges in a generic step of the operation and $V(q)$ the corresponding potential difference. The work to move a further dq from the negative to the positive plate is $dW = V(q)dq$, where $V(q) = q/C$. The total work required to charge the capacitor, namely the energy stored in the capacitor, is

$$W = U_E = \int_0^Q V(q)dq = \frac{1}{C} \int_0^Q qdq = \frac{1}{2} \frac{Q^2}{C}$$

The plates of a capacitor are always subject to an attractive force due to the electric field between them. Let us calculate this force in the simplest case of a parallel plate capacitor. Let A be the area of the plates, h their distance, and C the

capacitance. The charge of the capacitor is Q . Let us calculate the force F on one plate using the virtual works principle. The virtual work for a virtual change dh of the distance between the plates is equal to the energy variation, namely

$$dW = -Fdh = dU_E. \quad (3.11)$$

We must pay attention here to the fact that the virtual work is the *total* work to be done on the system from outside for the virtual variation. Does Eq. (3.11) contain all the work? The answer is yes, if we operate at constant charge, and no if we operate at constant potential difference. Indeed, in the second case, in addition to the work made by the force, work is needed to keep the potential difference constant. For that to happen, the charges on the plates must vary, and this requires electric work by the voltage source to which the plates are connected. When working at constant potential difference, we must include this electric work in the left-hand side of the equation. The final result is obviously the same in both cases. We proceed at constant charge and express the energy as a function of the charge, namely

$$U_E = \frac{1}{2} \frac{Q^2}{C} = \frac{Q^2 h}{2\epsilon_0 A}. \quad (3.12)$$

The force $F = -dU_E/dh$ is then

$$F = -\frac{Q^2}{2\epsilon_0 A} \quad (3.13)$$

where the minus sign indicates that the force is attractive. The corresponding pressure, namely force per unit area, is

$$\frac{F}{A} = \frac{\sigma^2}{2\epsilon_0}. \quad (3.14)$$

which is an expression we already know.

QUESTION Q 3.3. How much is the energy stored in a 20 μF capacitor charged to 220 V? \square

3.5 Energy in the Electrostatic Field

Let us go back to Eq. (3.7) and find an equivalent expression of the energy of a continuous charge distribution in terms of the electric field, rather than the potential and charge density. We start by eliminating the charge density from Eq. (3.7) using the Poisson equation $\rho = -\epsilon_0 \nabla^2 \phi$, namely

$$U_E = \frac{1}{2} \int_V \rho \phi dV = -\frac{\epsilon_0}{2} \int_{\text{all space}} \phi \nabla^2 \phi dV$$

where, for reasons that will become immediately clear, we have extended the integral to all the space profiting off the fact that $\rho = 0$ outside V . We can now have the field $\mathbf{E} = -\nabla\phi$ using the identity $\phi\nabla^2\phi = \nabla \cdot (\phi\nabla\phi) - (\nabla\phi)^2$. Substituting, we get

$$U_E = \frac{\epsilon_0}{2} \int_{\text{all space}} \nabla \cdot (\phi \mathbf{E}) dV + \frac{\epsilon_0}{2} \int_{\text{all space}} \mathbf{E}^2 dV$$

We evaluate the first integral by integrating on a spherical volume, say V_R , of finite radius R and then taking its limit for $R \rightarrow \infty$. Let Σ_R be the surface of the sphere and \mathbf{n} the unit vector outside normal to the surface. We use the Gauss divergence theorem and get

$$\lim_{R \rightarrow \infty} \int_{V_R} \nabla \cdot (\phi \mathbf{E}) dV = \lim_{R \rightarrow \infty} \int_{\Sigma_R} (\phi \mathbf{E}) \cdot \mathbf{n} d\Sigma$$

We assume the volume V to be limited in that in which $\rho \neq 0$, namely not to have charges at infinite distances, as is always the case in practice. Then, for R increasing at large distances outside V , the potential ϕ decreases (at least) as $1/R$ and E (at least) as $1/R^2$, while the surface Σ_R on which we integrate increases as R^2 . Consequently, the integral decreases as $1/R$ and its limit is zero. We are left with the second integral only and we finally have

$$U_E = \frac{\epsilon_0}{2} \int_{\text{all space}} \mathbf{E}^2(\mathbf{r}) dV. \quad (3.15)$$

This very important equation suggests that the electrostatic field energy is present everywhere in space where there is an electrostatic field, with an energy density, namely an energy per unit volume

$$w_E(\mathbf{r}) = \frac{\epsilon_0}{2} \mathbf{E}^2(\mathbf{r}). \quad (3.16)$$

We are induced to think that any infinitesimal space volume dV contains the energy $w dV = (\epsilon_0/2) E^2 dV$. The field energy pervades the entire space, even where there is no charge, namely ρ is null but \mathbf{E} is not. The energy density is larger where the field is larger, being proportional to its square. Note that energy is invariant under the rotations of the reference frame, namely it is a scalar. The simplest expression of a scalar in terms of a vector, like \mathbf{E} is, is just \mathbf{E}^2 .

QUESTION Q 3.4. At a distance of 1 cm from a uniformly charged straight wire, the field energy density is $w = 1 \text{ J/m}^3$. How much is the charge density of the wire? \square

Equations (3.7) and (3.15) are completely equivalent in electrostatics. Consequently, the existence of an energy density suggested by Eq. (3.15) cannot be verified through experiment. However, as we shall see in the subsequent chapters, under dynamical conditions, when charges and fields vary in time, Eq. (3.15) still holds, while Eq. (3.7) does not. Under dynamical conditions, Eq. (3.16) gets its complete physical meaning. Think, for example, of electromagnetic waves. A receiving device, like a cell phone or a TV set, detects the energy of an electric field that has been emitted by a source that is far away, where (moving) charges have generated that field. The field has, so to speak, detached itself from the sources and propagated carrying energy with it. When we talk of a vacuum, we mean space empty of matter. But such a vacuum it is not really empty; it contains energy (and linear and angular momentum too) of the field. We shall discuss these phenomena in Chap. 10.

Looking at Eqs. (3.7) and (3.15), one sees that the latter implies that energy of an electrostatic system is always positive, while for the former, it may be positive or negative. How can the two equations be equivalent? When we talk of energy, we always talk of an energy difference between the state under consideration and the state whose energy we have chosen to be of zero. The energy of a state is always the work to be done on the system to assemble that state starting from the zero-energy state. In the zero-energy state for Eq. (3.7), the point charges already exist, but do not interact because they are infinitely distant from one another. Equation (3.15) also includes the work that would be necessary to assemble the charges themselves. As we shall see in Sect. 3.6, this work is infinite in the case of the point charges.

As we already discussed in Chap. 2, when dealing with the field inside matter, we must distinguish the macroscopic field from the microscopic field. The microscopic electric fields are the fields existing inside the molecules and the atoms between electrons and nuclei, inside the nuclei between protons, and inside protons and neutrons between quarks. These fields are never static but vary rapidly in time. They are correctly described by quantum physics. They vary in space over distances that are very small compared to the macroscopic ones. The macroscopic field is an average of the microscopic field over distances and durations much larger and longer than the molecular sizes and variation times. The microscopic fields existing inside unperturbed microscopic objects, such as the nucleons, the nuclei, the atoms and the molecules, are already null at short distances outside them, at least as long as they are not perturbed. As a consequence, their contributions average to zero when considering the macroscopic field.

The situation is different when we consider the energies. Energy is proportional to the *square* of the field and the average of the square is not equal to the square of the average. The superposition principle does not hold. The squares of the microscopic fields do not average to zero. The corresponding energy is stored in matter as internal energy and mass. In our description of the macroscopic phenomena, when

considering the energy of a system, we shall always consider the work done to assemble it starting from molecules and atoms as given.

Let us give an example. Consider the conduction electrons in a metal. They move freely inside the metal like a gas. There are no walls on the surfaces but electrons are bound inside the metal by their interaction with the ions of the crystal. To extract an electron, we need to give it a certain energy, called work function. Inside the metal the macroscopic field is zero everywhere and so are the volume and surface charge densities. What is the mechanism keeping the conduction electrons inside? As a matter of fact, in the immediate neighboring of the surface, a fraction of the electrons does exit a bit, to extremely short distances from the surface. As a consequence, a positive charge layer forms at the surface and a negative one immediately outside it. The net surface charge density is zero, but between the two layers, an electric field exists. This is the field against which one must work to extract electrons from the metal. It is a microscopic field. The corresponding energy, namely the work function, is not included in the macroscopic energy.

We profit from the above description in noting that it suggests that the work function is a characteristic of the surface, not of the bulk of the metal (or of the conductor in general). It depends in particular on how the surface has been worked, on the presence of contaminants, water or oxides, etc.

We shall now discuss the energy of two configurations of capacitors as examples.

Energy of a parallel-plate capacitor.

Let A be the area of the plates of a parallel-plate capacitor, h their distance and Q its charge. We take the approximation that the field is uniform between the plates and zero immediately outside them (Fig. 3.3). In other words, we neglect the edge effects. In this approximation, the capacitance is $C = \epsilon_0 A/h$ and the energy is

$$U_E = \frac{1}{2} \frac{Q^2}{C} = \frac{Q^2 h}{2\epsilon_0 A}.$$

Let us check what Eq. (3.15) gives us. Recalling that $E = \sigma/\epsilon_0$ and that $\sigma = Q/A$, we have

$$U_E = \frac{\epsilon_0}{2} \int_{\text{all space}} \mathbf{E}^2 dV = \frac{\epsilon_0}{2} \int_{\text{all space}} \frac{Q^2}{\epsilon_0^2 A^2} dV = \frac{Q^2}{2\epsilon_0 A^2} Ah = \frac{Q^2 h}{2\epsilon_0 A},$$

which is equal to the above expression.

Fig. 3.3 The field of a parallel plate capacitor

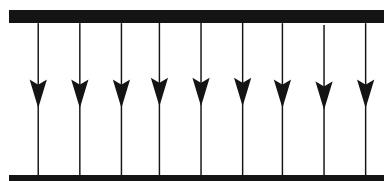
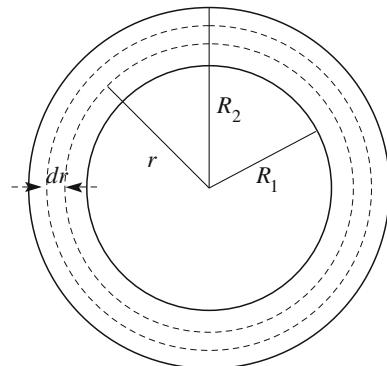


Fig. 3.4 Geometry of a spherical capacitor



Energy of a spherical capacitor.

Let R_1 and R_2 be the radii of the two spherical surfaces ($R_1 < R_2$), and let the capacitor be charged with the charge Q .

Recalling Eq. (2.20) for the capacitance, the energy of the capacitor is

$$U_E = \frac{1}{2} \frac{Q^2}{C} = \frac{1}{4\pi\epsilon_0} \left(\frac{1}{R_1} - \frac{1}{R_2} \right) Q^2.$$

On the other hand, the magnitude of the electric field in the capacitor at the distance r from the center is

$$E(r) = \frac{Q}{4\pi\epsilon_0 r^2}$$

To calculate the integral in Eq. (3.15), we take as dV the spherical shell between r and $r + dr$, as in Fig. 3.4. This volume is $dV = 4\pi r^2 dr$, and so we write

$$U_E = \frac{\epsilon_0}{2} \int_{R_1}^{R_2} \left(\frac{Q}{4\pi\epsilon_0} \right)^2 \frac{4\pi r^2}{r^4} dr = \frac{1}{2} \frac{Q^2}{4\pi\epsilon_0} \int_{R_1}^{R_2} \frac{dr}{r^2} = \frac{1}{2} \frac{Q^2}{4\pi\epsilon_0} \left(\frac{1}{R_1} - \frac{1}{R_2} \right)$$

giving us back the above equation.

3.6 The Energy of a Point Charge

In Sect. 3.1, we discussed the energy of a system of point charges and highlighted that the energy is purely interaction energy. We then analyzed continuous charge distributions and expressed their energy as energy of the electrostatic field. Let us now try to extend the latter point of view to point charges. Let q_1 and q_2 be two

point charges at the positions \mathbf{r}_1 and \mathbf{r}_2 , respectively, and let us calculate the energy of their field. If \mathbf{E}_1 and \mathbf{E}_2 are the fields of the two charges, the energy of the total field is

$$U_E = \frac{\epsilon_0}{2} \int \mathbf{E}^2 dV = \frac{\epsilon_0}{2} \int \mathbf{E}_1^2 dV + \frac{\epsilon_0}{2} \int \mathbf{E}_2^2 dV + \epsilon_0 \int \mathbf{E}_1 \cdot \mathbf{E}_2 dV.$$

One immediately sees that the first two terms on the right-hand side do not depend on the relative positions of the two charges. They also exist when the charges are infinitely distant and consequently do not interact. If there was only one charge, say q_1 , the energy of the field would be

$$U_E = \frac{\epsilon_0}{2} \int \mathbf{E}_1^2 dV$$

and similarly for q_2 . These are not interaction energies but proper energies of each charge. In Sect. 3.1, we did not include these terms, but only the interaction energy, which is the third term in the above expression.

The proper energies we have now found are a direct consequence of having associated the energy with the field, including the field of a single charge. Let us calculate the proper energy of the field \mathbf{E} of a point charge q . Let the charge be in the origin and let r be the distance from it. The energy in the infinitesimal spherical shell between r and $r + dr$, whose volume is $dV = 4\pi r^2 dr$, is

$$dU_E = \frac{\epsilon_0}{2} E^2 4\pi r^2 dr = \frac{\epsilon_0}{2} \left(\frac{q}{4\pi\epsilon_0 r^2} \right)^2 4\pi r^2 dr = \frac{q^2}{8\pi\epsilon_0} \frac{dr}{r^2}.$$

The total energy in the field is obtained by integrating up to infinity and down to zero, namely

$$U_E = \frac{q^2}{8\pi\epsilon_0} \int_0^\infty \frac{dr}{r^2} = \frac{q^2}{8\pi\epsilon_0} \left[\lim_{r \rightarrow \infty} \left(-\frac{1}{r} \right) - \lim_{r \rightarrow 0} \left(-\frac{1}{r} \right) \right]$$

The first limit is zero. The second limit, namely for the distance going down to zero is infinite. Hence, the energy of the electrostatic field of a point charge is infinite:

$$U_E = \frac{q^2}{8\pi\epsilon_0} \lim_{r \rightarrow 0} \left(\frac{1}{r} \right). \quad (3.17)$$

Let us try to understand the reason for that. This energy is the work needed to assemble a point charge. We shall start building a spherical shell of radius a by

bringing the charges from infinity (where they don't interact), and then squeezing the sphere, reducing its radius to zero. Let us calculate the necessary work.

To evaluate the work required to build the spherical surface, consider that we have already brought the charge q' onto the surface and express the work required to bring in an additional dq' . This is simply dq' times the potential at a of the existing charge, namely

$$dW = dU_E = \frac{q'}{4\pi\epsilon_0} \frac{1}{a} dq',$$

The total work is obtained by integrating from the initial value of the charge, namely 0, to the final one, which is q . We have

$$W = \frac{1}{2} \frac{q^2}{4\pi\epsilon_0} \frac{1}{a},$$

which is finite. To have a point charge we now need to squeeze down our spherical shell, reducing a to zero. The necessary work is

$$U_E = \frac{1}{2} \frac{q^2}{4\pi\epsilon_0} \lim_{r \rightarrow 0} \left(\frac{1}{a} \right), \quad (3.18)$$

which is just Eq. (3.17). The work, and consequently the energy in the field, is infinite because, as the charges approach one another, the repulsive forces between them increase without limit.

As a matter of fact, the same result is reached for any shape of the initial charge distribution. As another example, let us start from a solid sphere instead of from a spherical shell. Let us calculate the work required to build a sphere of charge q , radius a and uniform charge density ρ , bringing in charges from infinity. We build the sphere by bringing in infinitesimal spherical shells of charge. At the current step in the process, we have already built a sphere of radius r . The work to add the charge $dq = \rho 4\pi r^2 dr$ from infinity is given by this charge times the potential of the already existing charge. This charge is $q(r) = \rho(4/3)\pi r^3$. The work is then

$$dW = \phi(r) dq = \frac{q(r)}{4\pi\epsilon_0 r} \rho 4\pi r^2 dr.$$

The total work to be done is

$$W = \rho^2 \frac{4\pi}{3\epsilon_0} \int_0^a r^4 dr = \rho^2 \frac{4\pi}{15\epsilon_0} a^5.$$

Or, in terms of the final charge $q = (4/3)\pi a^3 \rho$

$$W = \frac{3}{5} \frac{q^2}{4\pi\epsilon_0} \frac{1}{a}.$$

The work to squeeze the sphere to zero radius, namely the energy of the field, is

$$U_E = \frac{3}{5} \frac{q^2}{4\pi\epsilon_0} \lim_{a \rightarrow 0} \left(\frac{1}{a} \right). \quad (3.19)$$

The limit is again infinite. The only difference with Eq. (3.18) is the factor of 3/5 instead of 1/2. The difference is relevant for finite radiiuses, but irrelevant for $a \rightarrow 0$. The divergence is the same. The same conclusion is reached for any shape of the charge distribution with which one might start. The work required to squeeze it to zero diameter is, in any case, infinite.

The infinite energy of the field of the point charges has no practical consequence, because the point charges, such as, for example, electrons, are given. We can neither build them nor destroy them. A serious theoretical difficulty exists, however. The concept of energy of the field, which is essential in the development of electromagnetism, is incompatible with point charges.

Summary

In this chapter, we have studied the energy of systems of charges at rest in a vacuum and on conductors. We have learned the following important concepts:

1. The expression for the energy of a system of charges at rest in given positions.
2. The expression for the energy of a system of conductors at rest in given positions.
3. The energy of a capacitor.
4. The energy and energy density of the electrostatic field.
5. The infinite energy of the field of a point charge.

Problems

- 3.1. Five conductors with charges $Q_1 > Q_2 > Q_3 > Q_4 > Q_5$ are located far from one another. We move one at a time until they are all near to one another. We can proceed in the order 1, 2, 3, 4, 5, or 5, 4, 3, 2, 1. How do the total necessary works compare?

- 3.2. A point charge $q = 1 \text{ pC}$ is in a position in an electrostatic field generated by fixed charges of unknown positions and values. We, however, know that the potential of that point relative to infinity is 1000 V. How much work should be done to move the charge to infinity?
- 3.3. We charge a parallel plate capacitor by connecting its plates to the poles of a battery, and then disconnect the plates. Is any work needed to double the distance between the plates? If so, where does that work go?
- 3.4. We have five identical capacitors. We connect two of them in parallel and two in series. Then, we separately connect to the poles of a battery (a) the unique capacitor, (b) the pair in parallel, and c) the pair in series. In which case is the energy delivered by the battery the largest?
- 3.5. A proton and an alpha particle ($Z = 2$) move from the negative to the positive plate of a parallel plate capacitor starting from rest. What is the ratio of their final kinetic energies?
- 3.6. We want to increase the distance between the plates of a parallel plate capacitor of area $A = 50 \text{ cm}^2$ by 0.1 mm. The charge is $Q = 20 \text{ nC}$. Neglecting the edge effects, how much work is necessary, working at constant charge?
- 3.7. A charge $Q = 10 \text{ nC}$ is uniformly distributed along the surface of a sphere of radius $R = 1 \text{ cm}$. How much is the energy of the electrostatic field? How much is the energy of the field if the same charge is uniformly distributed in the volume of the sphere? In the latter situation, what is the fraction of the field energy outside the sphere?
- 3.8. Consider a spherical metal shell of radius R . There is a small hole in the shell. If there is no charge on the shell, how much work is needed to bring a charge q from infinity through the hole to the center of the sphere? How much if the sphere has a charge Q ?

Chapter 4

Dielectrics

Abstract In this chapter we discuss the electric properties of the insulators, or dielectrics, as these materials are also called. There are no free charge carriers in dielectrics, which consequently do not conduct electricity. However, dielectrics react in response to an electric field. We introduce the concept of the dielectric constant, find the differential equations of the electric field for normal dielectrics, and solve those equations. We give a microscopic interpretation of the described phenomena. We discuss the electrostatic energy in the presence of a dielectric medium.

In this chapter, we describe the properties of the static electric field in the presence of non-conducting media, namely insulators. These materials also began to be referred to as dielectrics when it was discovered that they are penetrated by the electric field. Dielectrics differ from conductors because they do not contain free charge carriers. However, something happens when they are in an electric field, namely the atomic charges rearrange and reorient under its action. These displacements are extremely small, even on the atomic scale, but they have relevant and macroscopically observable consequences. We are talking about dielectric polarization, a phenomenon in which every elementary volume of the medium acquires a non-zero electric dipole moment.

After having introduced the concept of the dielectric constant in Sect. 4.1, we start by describing and interpreting polarization in the simplest case of a uniform electric field in Sect. 4.2. We then go to the general case in the subsequent sections. In Sect. 4.6, we find the differential equations of the electric field valid for a large class of dielectric media, the so-called normal dielectrics, and we solve those equations. We introduce the “electric displacement” field, which is a useful auxiliary field, with a misleading name of historical origin. As a matter of fact, it does not displace anything. Its use simplifies the solution of the field equations.

In Sect. 4.7, we try to give a microscopic interpretation of the just-described phenomena. The interpretation will necessarily be quite approximate because, at this dimensions, we should give a quantum description. However, it will suffice to understand the essence of the mechanisms. We shall see that two types of dielectric materials exist, depending on their molecules. The molecules of polar dielectrics have a non-zero intrinsic dipole moment, namely also existing in the absence of an electric field, while the molecules of the nonpolar dielectrics have null intrinsic dipole moment.

The action of an applied electric field on the polar molecules is a tendency to orient them in its direction. The resulting effect is statistical equilibrium with the contrasting action of thermal agitation. The action of the applied field on the nonpolar molecules is to deform their charge distribution with the effect of inducing a non-zero dipole moment. The effects of the electric field on the polar materials are, in general, larger than those on the nonpolar ones. Particularly large is the dielectric constant of water, whose molecule, as we have seen in Chap. 1, has quite a large dipole moment. This fact is the cause of the hydrolysis of salts, a phenomenon on which life itself depends.

In Sect. 4.8, we analyze the electric field and the electric displacement in cavities inside a dielectric and see how they are different for different shapes of the cavity. We shall also see that the electric field acting on the generic molecule in a medium is not exactly the applied field. This is because the former also receives the contributions of the dipole moments induced in the other molecules.

In Sect. 4.9, we study the energy of a system of free charges on conductors immersed in a fluid dielectric medium and how it can be thought of, in this case too, as field energy.

4.1 Dielectric Constant

In the previous chapter, we discussed the properties of conductors. Here, we discuss the electric properties of another class of materials, namely insulators, or dielectrics, as they are also called. There are no free charge carriers in dielectrics, which consequently do not conduct electricity. However, dielectrics react in response to an electric field. Even if their charges cannot move over large distances, they do so on the sub-molecular scale. The local rearrangement of molecules or atoms in response to an electric field produces observable and important effects.

Historically, the question of possible electric effects in materials that were used, at the time, for insulating purposes was raised for the first time by Michael Faraday. In 1837, he performed a series of experiments with two equal spherical condensers, one with air and one with different insulators in turn filling the gap between the metallic surfaces. He discovered the capacities to be larger in the presence than in the absence of insulators, by factors depending on the material. After the discovery, insulators were also called dielectrics, from the Greek dia (=through)-electricity.

We shall describe the relevant observations in a simplified form. Let us consider a parallel plate capacitor made of two metallic plane plates of a few dozen centimeters in diameter at a few centimeters distance. In principle, there should be a vacuum between the plates, but in practice, we can work in air, which differs from a vacuum by one half of a percentage for what we are going to discuss (see Table 4.4). Let A be the area of the plates and d their distance. Its capacitance is

$$C_0 = \varepsilon_0 A/d \quad (4.1)$$

where the subscript 0 stands for it being in a vacuum. Let us connect the capacitance to a battery of potential difference V_0 and then disconnect it. The charge of the capacitor is $Q = V_0/C_0$. Let us now introduce between the plates a dielectric slab, made, for example, of PVC, of thickness d filling the space between the plates of the capacitor. If we now measure the potential difference, we find a value V , which is less than V_0 . If we repeat the process several times, starting from different initial potential differences V_0 in air, we find values of V which are always proportional to V_0 , namely that $V = V_0/\kappa$. The constant κ is a characteristic of the medium and is called the dielectric constant of the material.

The observations we just made tell us that the system is still a capacitor, namely there is complete induction between the plates and the potential difference is proportional to the charge of the plates. The capacitance is, however, now larger, being

$$C = \kappa C_0 = \kappa \epsilon_0 A/d = \epsilon A/d \quad (4.2)$$

where we introduced the new constant

$$\epsilon = \kappa \epsilon_0, \quad (4.3)$$

which is called the *permittivity of the material* (indeed, ϵ_0 is the vacuum permittivity).

Let us now see the physical reasons for these observations. The fact that the potential difference is diminished by the presence of the dielectric, while the distance between plates does not vary, implies that the electric field diminishes by the same ratio, namely that

$$E = E_0/\kappa. \quad (4.4)$$

Let us consider the system as represented in Fig. 4.1 and apply the Gauss law to the surface S , which is a small cylinder with one face inside the metal plate and the other face inside the dielectric. The figure shows an enlarged version of the very thin air layers between the plates and the dielectric, where the field has its value in vacuum E_0 . In the dielectric, the field is E .

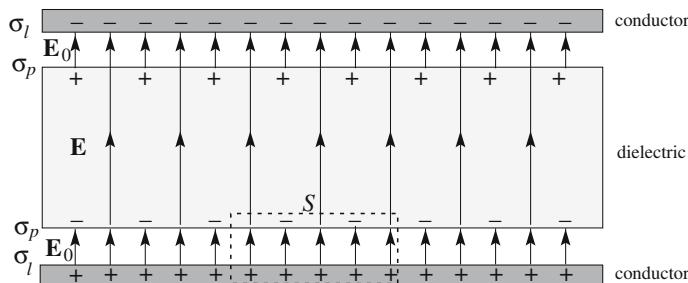


Fig. 4.1 A parallel plate capacitor with a dielectric filling its gap. The thickness of the layers between the plates and the dielectric is exaggerated

The flux of the field outgoing from S is $E \cdot S$. Indeed, the field is null in the conductor and parallel to the lateral surface of the cylinder. Consequently, only the face of the cylinder in the dielectric contributes to the flux. Let us call the charge on the surface of the conductor “free”, it being free of movement, and let us indicate its density with σ_f . This charge did not change when we introduced the dielectric slab. On the other hand, the flux $E \cdot S$ is equal to the charge inside S divided by ϵ_0 , namely it is less than it is in absence of the dielectric. We must conclude that there are charges on the surface of the dielectric. We call these *polarization charges* (for reasons that will become clear soon) and indicate their surface density with σ_P . Expressed in a formula, these conclusions are

$$S \cdot E = \frac{|\sigma_f| - |\sigma_P|}{\epsilon_0} S$$

namely

$$E = \frac{|\sigma_f| - |\sigma_P|}{\epsilon_0}, \quad (4.5)$$

while obviously the field in the vacuum is

$$E_0 = \frac{|\sigma_f|}{\epsilon_0}. \quad (4.6)$$

4.2 Polarization of a Dielectric

To progress in our analysis, we need to think, at least in a first approximation, as to what happens inside a dielectric. We shall give more details in Sect. 4.7, where we shall see that two situations exist, depending on the molecules having or not having a permanent dipole moment. For the moment, it is enough to consider monoatomic molecules without a permanent dipole moment. We can think of the atoms as being spherical distributions of negative charge (the electrons) with a positive equal and opposite point charge in the center. The centers of the negative and positive charges coincide and the dipole moment is zero. This is true as long as there is no external field. If we now switch on the field (namely we move the dielectric in the capacitance), the electron cloud and the nucleus become subject to opposite forces, which tend to separate the two centers. The atom reacts with a force proportional to the deformation. Let δ be the position vector at equilibrium taken from the negative to the positive charge centers. If q is the charge of the nucleus (and the opposite charge of the electrons), the deformed atom has acquired a dipole moment

$$\mathbf{p} = q\boldsymbol{\delta}. \quad (4.7)$$

If the medium is isotropic, \mathbf{p} has the same direction and sense as the field. If the electric field is not extremely strong, as is usually the case, the deformation $\boldsymbol{\delta}$ is very small and is proportional to the field. Hence, the induced dipole moment is proportional to the field as well. We speak of *linear conditions* and also of a *linear dielectric*. If a dielectric is linear, isotropic and homogeneous, it is said to be a *normal dielectric*. We now introduce the constant α , called the *atomic polarizability constant*. We have

$$\mathbf{p} = \alpha \mathbf{E} \quad (4.8)$$

The dimensions of α are $C^2 m/N$. The process we have described is called electrostatic polarization. The constant α is characteristic of the material. If it is larger, the material is “easier” to polarize.

The corresponding macroscopic quantity is the electric dipole moment per unit volume, which we indicate with \mathbf{P} and which is called the *polarization density* or simply the *polarization*. If n_p is the number of atoms per unit volume, we have

$$\mathbf{P} = n_p \mathbf{p} = n_p q \boldsymbol{\delta} \quad (4.9)$$

The polarization vector \mathbf{P} is of main importance in the study of dielectrics, as we shall see in this chapter and in the subsequent ones. It is a vector field, which can be uniform or not. Under the conditions we are considering, \mathbf{P} is proportional to the field, namely

$$\mathbf{P} = \alpha n_p \mathbf{E}. \quad (4.10)$$

This relation is often expressed in terms of still another constant, the electric susceptibility χ_e , as

$$\mathbf{P} = \epsilon_0 \chi_e \mathbf{E}. \quad (4.11)$$

The electric susceptibility is characteristic of the material, and, in the case we are discussing, is obviously $\chi_e = n_p \alpha / \epsilon_0$.

The measurement unit of \mathbf{P} , which is a distance times a charge divided by a volume, is $C m^{-2}$. Recalling the dimensions of ϵ_0 , we see that α / ϵ_0 has the dimensions of a volume and that χ_e is dimensionless. As we shall see, α / ϵ_0 has the order of magnitude of the atomic volumes.

4.3 Uniform Polarization

Consider again a dielectric medium between the plates of a parallel plate capacitor, neglecting the edge effects. This is the simplest situation to analyze, the electric field \mathbf{E} being uniform on the entire dielectric. Let us think of the unperturbed medium, namely in the absence of an electric field, as the overlap of two uniform,

equal and opposite, charge distributions. The resulting charge density is zero everywhere. If we now turn on the field, the charge distributions shift, the positive one moving in the direction of the field, the negative in the opposite direction.

Let δ be the relative shift, as shown in Fig. 4.2. The total charge density is still zero inside the medium, but it is not so in two layers of thickness δ near the two faces. This is the surface polarization density σ_P . The “surface” is not a geometrical but a physical one. Its thickness is very small, smaller than the picometers. σ_P is the charge per unit surface, meaning the charge in a volume of $1 \times \delta$. If n_p is the number per unit volume of charges and q their value, then we have $|\sigma_P| = n_p q \delta$. Recalling Eq. (4.9), we conclude that

$$|\mathbf{P}| = |\sigma_P| \quad (4.12)$$

In conclusion, if the polarization is uniform and normal to the surface, the surface polarization density is equal to the magnitude of polarization in the medium. We can now write Eq. (4.5) as $E = |\sigma_f|/\epsilon_0 - P/\epsilon_0$, and also

$$E + P/\epsilon_0 = |\sigma_f|/\epsilon_0$$

This equation can be solved if we know the relation between \mathbf{E} , which is the cause, and \mathbf{P} , which is the effect. Under linear conditions Eq. (4.11), we have

$$|\mathbf{E}|(1 + \chi_e) = |\sigma_f|/\epsilon_0.$$

Notice that the right-hand side of this equation is just the electric field in a vacuum, showing that E is proportional to E_0 . Considering that $\kappa E = E_0$, the equation establishes a relation between susceptibility and the dielectric constant, namely

$$\kappa = 1 + \chi_e. \quad (4.13)$$

Note that susceptibility, as defined by Eq. (4.11), is a positive quantity, because \mathbf{P} has the same sense as \mathbf{E} . Consequently, the dielectric constant is $\kappa > 1$, corresponding to the observation that the electric field in a dielectric is smaller than that in a vacuum.

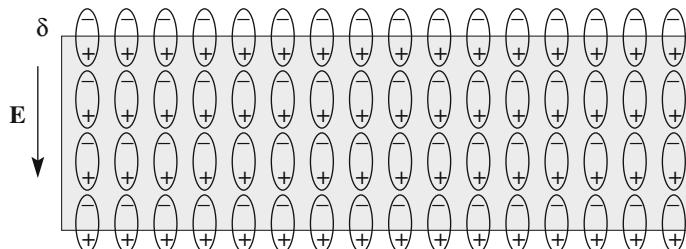


Fig. 4.2 Polarization of a dielectric in a uniform field

4.4 Non-uniform Polarization

If the electric field \mathbf{E} and the polarization \mathbf{P} are uniform, as we have just seen, polarization charges appear on the dielectric surfaces, but its volume remains globally neutral. Let us analyze the reason for that.

Let us consider a thin layer in the dielectric of thickness d between two surfaces, say S_1 and S_2 , parallel to the external faces of the dielectric. In absence of a field, the net charge in the layer is zero. The relevant consequence of turning on the field is the displacement of the charges of one sign relative to the other. We can think, for example, of the negative charges remaining in place and the positive ones moving. Let δ_1 and δ_2 be their displacements, at S_1 and S_2 , respectively, where we have considered that the two displacements might be different. As shown in Fig. 4.3, this means that all the positive charges that are located at a distance smaller or equal to δ_1 above S_1 enter (contributing to an increase in its charge), while the positive charges within a distance δ_2 above S_2 exit the layer (diminishing its charge). If the two displacements are equal, namely if $\delta_1 = \delta_2$, the net charge density in the layer remains zero. The displacement being proportional to the electric field, this is the case if $\mathbf{E}_1 = \mathbf{E}_2$. This is why the polarization volume charge density is zero in a uniform electric field.

However, when the field is not uniform, namely if $\mathbf{E}_1 \neq \mathbf{E}_2$, then it is also $\delta_1 \neq \delta_2$, and a volume charge density may appear.

Before analyzing this case, we consider the case of a uniform field whose direction is not perpendicular to the surfaces of the dielectric. Let θ be the angle between the electric field and normal to the surfaces, as in Fig. 4.4. The only difference between this and the situation in the last section is that now the shift δ is at an angle with the surfaces. Consequently, the thickness of the polarization charge “surface” is $d \cos \theta$, rather than δ .

Fig. 4.3 A dielectric layer and the charge displacements induced by an electric field

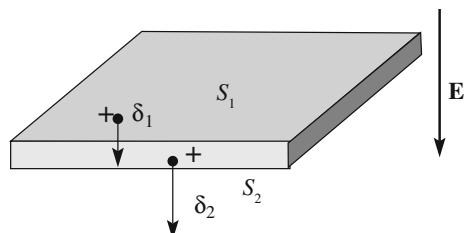
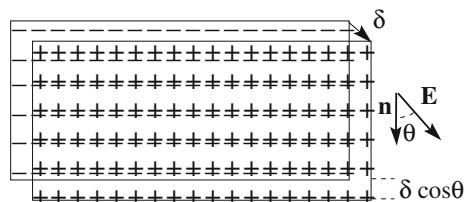


Fig. 4.4 Polarization of a dielectric in a uniform electric field at an angle with the surfaces



Now, we have $|\sigma_P| = n_p q \cos \theta = P \cos \theta$, and, indicating the unit vector of the outgoing normal with \mathbf{n} ,

$$\sigma_P = \mathbf{P} \cdot \mathbf{n}. \quad (4.14)$$

This is the generalization of Eq. (4.12) to a uniform electric field of arbitrary direction with the dielectric surface.

Let us now give a slightly different interpretation of Eq. (4.14). As a matter of fact, any unitary surface, even if *inside* the dielectric, when the electric field goes on is crossed by a charge equal to $\mathbf{P} \cdot \mathbf{n}$, where \mathbf{n} is its unit normal vector. As we have seen, this does not give rise to a volume charge density if \mathbf{P} is uniform, but it does when \mathbf{P} is a function of position.

Let V be an arbitrary volume inside a dielectric and Σ its surface. Let \mathbf{n} be the outgoing normal unit vector of Σ (see Fig. 4.5). The total charge in V is zero in absence of a field. When the field is turned on, the dielectric polarizes, with a polarization generally different at different points. The displacements of the positive charges (always relative to the negative ones) may bring them in or out of the volume through the Σ . The elementary surface $d\Sigma$ is crossed, in the sense of \mathbf{n} , by the charge $dQ = \mathbf{P} \cdot \mathbf{n} d\Sigma$ (which obviously may have either sign). The process gives origin to the appearance in V of a net charge ΔQ equal and opposite to the net charge that went out from Σ , namely

$$\Delta Q = - \int_{\Sigma} \mathbf{P} \cdot \mathbf{n} d\Sigma.$$

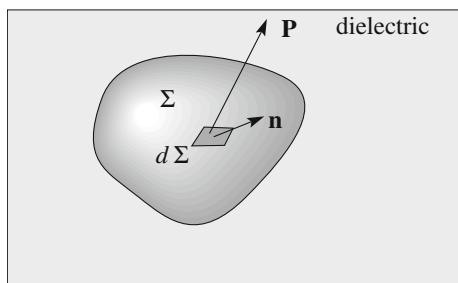
This charge is distributed in V with a certain volume charge density, which we call ρ_P , and write as

$$\Delta Q = \int_V \rho_P dV.$$

Using the Gauss divergence theorem, we have

$$\Delta Q = - \int_{\Sigma} \mathbf{P} \cdot \mathbf{n} d\Sigma = - \int_V \nabla \cdot \mathbf{P} \cdot \mathbf{n} dV = \int_V \rho_P dV.$$

Fig. 4.5 A surface *inside* a dielectric, and the polarization and *outside* normal in a surface element



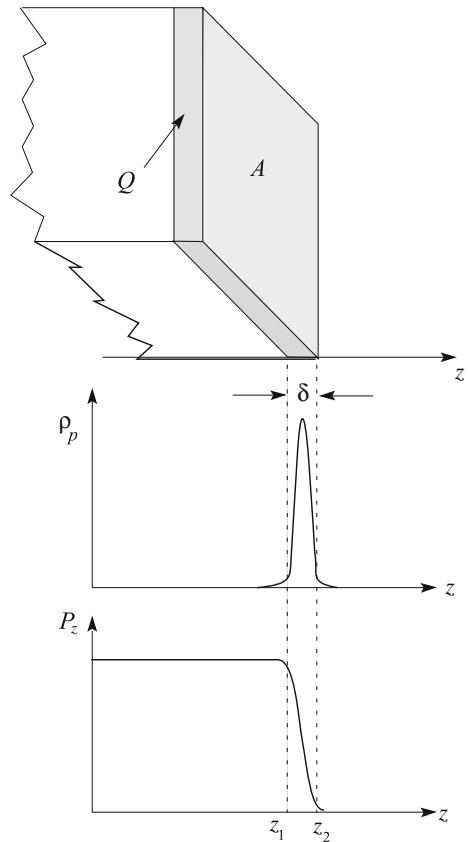
The last equality holds for an arbitrary volume. Hence, the integrands must be equal, and we have

$$\rho_P = -\nabla \cdot \mathbf{P}. \quad (4.15)$$

We notice the following in regard to this important equation. First, what we called ρ_P is physically just a charge density like any other one; the subscript is simply there to recall the physical process that gives it its origin. Secondly, note that when \mathbf{P} is uniform, its divergence is zero everywhere and we again find back that the volume polarization charge density is zero. However, we now see that this is a sufficient, but not necessary condition for having $\rho_P = 0$. For that, it is enough that \mathbf{P} is solenoidal, namely if its divergence is everywhere zero.

We shall now confirm that our results agree with what we found in Sect. 4.4. Consider a uniformly polarized dielectric with the surfaces perpendicular to the field, as in Sect. 4.4, in a region near to its surface. Let z be an axis parallel to the electric field. As we said, the surface charge density σ_P occupies a volume of a very small, but not zero, thickness δ , which is shown magnified in Fig. 4.6. We can think

Fig. 4.6 Relation between surface and volume density for uniform polarization



in an equivalent manner to a volume charge density, say ρ_P , which is everywhere zero, but in the layer between z_1 and z_2 at the surface ($z_2 - z_1 = \delta$). In this layer, ρ_P is very large, because δ is very small. We find the relation between ρ_P and σ_P , thinking that

$$\sigma_P = \int_{z_1}^{z_2} \rho_P(z) dz = \langle \rho_P \rangle \delta.$$

The polarization vector \mathbf{P} has its uniform internal value up to z_1 and is zero at z_2 . Taking into account that the only component not identically zero is P_z , Eq. (4.15) tells us

$$\rho_P = -\nabla \cdot \mathbf{P} = -\frac{\partial P_z}{\partial z} = \frac{P(z_1) - P(z_2)}{\delta} = \frac{P}{\delta} = \frac{\sigma_P}{\delta}.$$

We see that the two descriptions agree.

4.5 Electrostatic Equations in a Dielectric

The differential equations for the electric field are

$$\nabla \cdot \mathbf{E} = \rho / \epsilon_0 \quad (4.16)$$

and

$$\nabla \times \mathbf{E} = 0. \quad (4.17)$$

The equations hold, under static conditions, both in a vacuum and inside matter, and inside a dielectric in particular. The important difference is that, in a dielectric, the charge density in Eq. (4.16) that gives origin to the field does not only consist of the charges that, so to speak, are under our control, but also of those induced in the dielectric through the polarization phenomenon, which is unknown. We shall now look for expressions of the field solely as a function of the charges we know, polarization charges excluded.

Let us consider a region of space full of a fluid dielectric, for example, an oil, and let us arrange in this region a number of metallic conductors, the shapes and positions of which we have chosen. Let us give to each conductor a certain charge that we know. We shall call them *free charges*, because each charge is free to move on its conductor, and indicate its density with ρ_f . In the absence of the dielectric, we could find the electric field by solving Eqs. (4.16) and (4.17), which are equivalent to the Laplace equation for the potential, with the boundary conditions we have defined with our conductors. In the presence of the dielectric, the polarization

charges contribute to the field too. Let ρ_P be the polarization charge density. Equation (4.16) holds that $\nabla \cdot \mathbf{E} = \rho_f/\epsilon_0 + \rho_P/\epsilon_0$. The second term on the right-hand side is unknown. Let us eliminate it using Eq. (4.15) and move it to the left-hand side, obtaining

$$\nabla \cdot (\mathbf{E} + \mathbf{P}/\epsilon_0) = \rho_f/\epsilon_0. \quad (4.18)$$

This expression suggests that we define the auxiliary field as

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}. \quad (4.19)$$

For historical reasons, we did not choose $\mathbf{D} = \mathbf{E} + \mathbf{P}/\epsilon_0$, which looks more natural. \mathbf{D} is called the *electric displacement* field, another misleading name, because it does not measure any displacement, that came about for historical reasons. The field has the same physical dimensions as the polarization density and is consequently measured in $C m^{-2}$. The vector field \mathbf{D} satisfies the differential equation

$$\nabla \cdot \mathbf{D} = \rho_f. \quad (4.20)$$

This means that the sources of \mathbf{D} are the free charges alone (under static conditions). The field lines of \mathbf{D} exit from and enter into points at which free charges are located. The lines of \mathbf{D} cross the polarization charges, for example, the charges on the surface of the dielectric plate considered in Sect. 4.1, as if they did not exist.

Up to now, we have not solved our problem at all. We have just hidden an unknown in the left-hand side of the equation. We need further information, namely the relation between \mathbf{D} and \mathbf{E} , or, equivalently, between \mathbf{P} and \mathbf{E} . This relation describes a situation that, in principle, is very complex, namely how matter reacts to the field, producing the polarization, a phenomenon that implies an enormous number of charges. In practice, in the vast majority of situations, we can consider the dielectric to be normal and Eq. (4.11) to hold. We have

$$\mathbf{D} = \epsilon_0(1 + \chi_e)\mathbf{E} = \epsilon_0\kappa\mathbf{E} = \epsilon\mathbf{E} \quad (4.21)$$

and for the electric field

$$\nabla \cdot (\kappa\mathbf{E}) = \rho_f/\epsilon_0. \quad (4.22)$$

The curl of the electric field obeys the same equation in a dielectric as it does in a vacuum, namely $\nabla \times \mathbf{E} = 0$. We now have the divergence of $\kappa\mathbf{E}$ and the curl of \mathbf{E} , which are two different fields. Things become simple for a normal dielectric, where κ is uniform and we can move it into the curl, obtaining

$$\nabla \times (\kappa\mathbf{E}) = 0. \quad (4.23)$$

We see that the same charge distribution in a vacuum produces the electric field \mathbf{E}_0 , satisfying the equations $\nabla \cdot \mathbf{E}_0 = \rho_f/\epsilon_0$ and $\nabla \times \mathbf{E}_0 = 0$, which are the same equations. If the boundary conditions are also the same, in particular, if the dielectric is infinitely extended, or more realistically, if its surfaces extend far enough, the solution must be the same as well, namely

$$\kappa \mathbf{E} = \mathbf{E}_0. \quad (4.24)$$

In a normal and infinite dielectric medium, the electric field generated by a certain distribution of free charges is equal to the field generated in a vacuum by the same distribution with the difference that its magnitude is smaller everywhere by the same factor κ . The complex actions of the polarization charges end up in a very simple result.

Suppose now that we repeat the Coulomb experiment in an infinitely extended fluid that is a normal dielectric of constant κ . Let q_1 and q_2 be the free charges, namely the charges on the two small spheres of the experiment. In the presence of the dielectric, the field everywhere is κ times smaller than in vacuum, and, is consequently, so is the force between the free charges. We can conclude that the force between two point charges at rest in a dielectric of the mentioned characteristics is

$$F = \frac{1}{4\pi\epsilon} \frac{q_1 q_2}{r^2}. \quad (4.25)$$

where $\epsilon = \kappa\epsilon_0$. As a matter of fact, this is the usual situation when we do our experiments in air, which is a dielectric. The effects are extremely small because the dielectric constant of air at STP is $\kappa = 1.0005$.

Equation (4.25) is known as the Coulomb law in a dielectric. As opposed to the Coulomb law in vacuum, it is not a universal law, being valid only under the assumptions we have made for the dielectric.

QUESTION Q 4.1. Two point charges at rest at 50 cm distance interact with a certain force in a vacuum. By which factor should their distance be changed if they are immersed in oil with dielectric constant $\kappa = 5$ to have the same force? \square

We now show that the polarization volume charge density inside a normal dielectric is zero. We start by finding the relation between \mathbf{P} and \mathbf{D} using Eqs. (4.11) and (4.21), namely

$$\mathbf{P} = \frac{\kappa - 1}{\kappa} \mathbf{D}. \quad (4.26)$$

The dielectric being normal, κ is independent of position and we can write

$$\rho_P = -\nabla \cdot \mathbf{P} = -\nabla \cdot \left(\frac{\kappa - 1}{\kappa} \mathbf{D} \right) = -\frac{\kappa - 1}{\kappa} \nabla \cdot \mathbf{D}.$$

But the divergence of \mathbf{D} is zero where there are no free charges, hence, inside a normal dielectric, $\rho_P = 0$, which does not exclude polarization charges on the surfaces of the dielectric.

QUESTION Q 4.2. The magnitude of the electric displacement in a parallel plate capacitor is $D = 10 \mu\text{C}/\text{m}^2$. What is the charge density on the plates? \square

Let us consider an example useful for understanding why the electric field in a dielectric is reduced by the factor κ . Consider a small conducting sphere with radius R and charge $+Q_f$ immersed in a normal fluid dielectric of very large extension, as shown in Fig. 4.7.

As we have just seen, there are no polarization charges in the volume. However, there is a negative surface density on the surface of the dielectric touching the sphere. Let Q_P be this charge. The corresponding positive charges are far away, on the free surface of the dielectric. We can neglect their effects.

The field at an arbitrary point P of the dielectric is equal to the field of a point charge $Q_f + Q_P$ in the center of the sphere. The charge “seen” by P is then smaller (Q_P is negative) than in vacuum under the same conditions. We say that polarization partially shields the free charge.

Let us see how the situation changes as a function of the radius of the sphere. When R decreases, the field of Q_f on the spherical surface of the dielectric increases by R^{-2} . The surface density increases by the same factor. However, Q_P does not vary because the surface area decreases by R^{-2} .

Before concluding, let us look at a simple device, called the electric pendulum, used to show the presence of an electric field. It consists of a small piece of a light dielectric, cork, for example, attached to a wire, as shown in Fig. 4.8. It is attracted, for example, by a piece of plastic or a comb that we have rubbed. Notice that the attraction force is not only due to the presence of an electric field but also to the fact that the field is not uniform.

The electric field induces in the pendulum body a polarization \mathbf{P} parallel and in the sense of the field. The pendulum becomes an electric dipole. The resultant force on it is proportional to the induced dipole moment, which is proportional to the

Fig. 4.7 A positive charged sphere in a normal dielectric medium

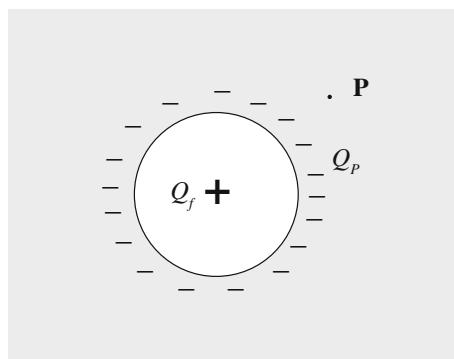
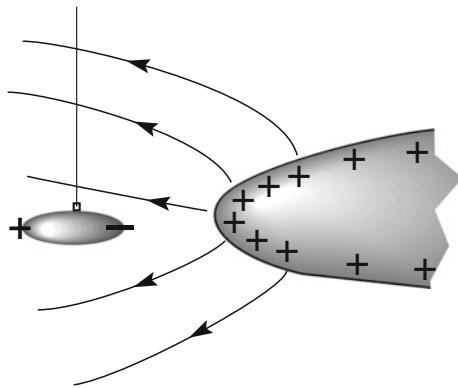


Fig. 4.8 Electric pendulum in an inhomogeneous field



field, and in addition to the gradient of the field. Note that the force is always an attraction toward the region of higher field. This is because the induced \mathbf{P} has the same direction and sense as \mathbf{E} .

4.6 Linear and Isotropic Dielectrics

In this section, we discuss the limits within which Eq. (4.11) holds. The equation gives a relation between the effect, which is the polarization \mathbf{P} , and its cause, which is the field \mathbf{E} . It contains two pieces of information: (a) the magnitude of the effect is proportional to the magnitude of the cause, and (b) the direction of \mathbf{P} is equal to the direction of \mathbf{E} . Let us consider the two points separately.

Point (a) is logically an approximation, which, in practice, is almost always very good. Indeed, we can say that, under time independent conditions, a certain functional relation between the magnitudes of the two vectors should exist, say $P = P(E)$. We do not know this function, but we can expand it in a Taylor series and, for values of E not too large, stop at the first term, namely

$$P(E) = P(0) + \left(\frac{dP}{dE}\right)_0 E + \dots = \left(\frac{dP}{dE}\right)_0 E + \dots$$

where we took into account that there is no polarization in the absence of an applied electric field, namely that $P(0) = 0$. This is the proportionality relation. Clearly, only experiments can tell us what “values of E not too large” means. However, in practice, the macroscopic electric field we are able to produce is always much smaller than the field inside the atoms. We shall see an example of that in Sect. 4.7. Consequently, the atomic deformations we are able to induce are always very small and the linear approximation is good.

Let us now consider point (b), namely the vector nature of the problem. In order for \mathbf{P} to be parallel to \mathbf{E} , the medium must be isotropic, namely its reaction to the

field should not depend on the direction. As a counterexample, consider a crystal lattice. We can schematically think of the structure in Fig. 4.9, which is composed of a lattice of square rigid frames. Four springs attached to the four sides of each mesh hold a sphere in the center. Let x and y be the directions of the sides. The two springs in one direction are equal, but different from the other pair. Let K_1 and K_2 be the spring constants of the x and y pairs, respectively.

Let all the spheres have the same charge Q and let us apply an electric field \mathbf{E} . If \mathbf{E} is in the x direction, the displacement δ_x of each sphere will be (for small displacements) in the x direction too, and given by the relation $2K_1\delta_x = QE$, where the factor 2 is the result of there being two springs. Similarly, if \mathbf{E} is in the y direction, the displacements δ_y are given by $2K_2\delta_y = QE$. When the direction of \mathbf{E} is arbitrary, with components E_x and E_y , the components of the displacement are

$$\delta_x = \frac{QE_x}{2K_1}, \quad \delta_y = \frac{QE_y}{2K_2}.$$

Hence, if $K_1 \neq K_2$, the direction of the displacement $\boldsymbol{\delta} = (\delta_x, \delta_y)$ is different from the direction of \mathbf{E} . The angles of \mathbf{E} and $\boldsymbol{\delta}$ with the x -axis are, respectively, $\alpha = \arctan(E_y/E_x)$ and $\beta = \arctan(K_1E_y/K_2E_x)$. As an example, shown in Fig. 4.10, suppose that $\alpha = 45^\circ$, namely that $E_y = E_x$ and $K_2 = 2K_1$. Then, $\beta = \arctan(1/2) = 26.6^\circ$.

Notice that the structure in Fig. 4.9 is anisotropic, both at the microscopic level, namely at the level of the cell, and at the macroscopic one, namely at the level of the array. A solid body can be thought to be made of a very large number of cells. If

Fig. 4.9 A lattice of springs and spheres representing a crystal

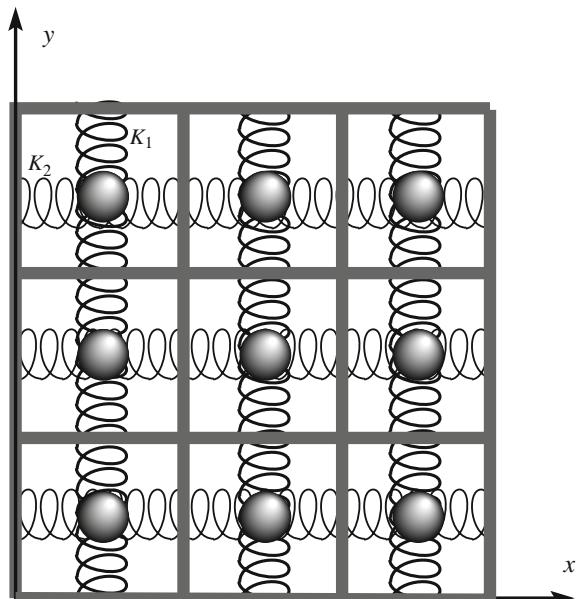
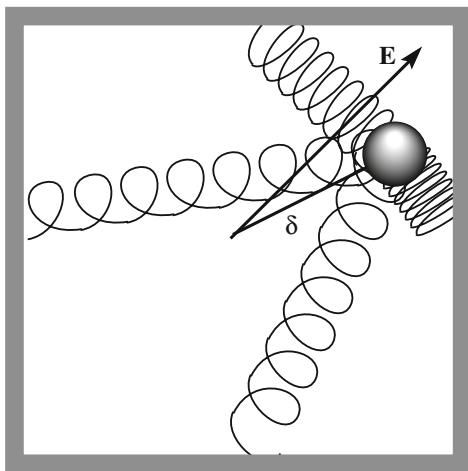


Fig. 4.10 A cell, field \mathbf{E} at 45° and displacement δ at 26.6°



the body is a single macroscopic crystal, all the displacements are in the same direction and consequently the direction of \mathbf{P} may not be parallel to \mathbf{E} . In the case of the polycrystalline media, such as metals, the cells are arranged in the same direction in each microcrystal, but the relative orientations between microcrystals are random. The components of \mathbf{P} perpendicular to \mathbf{E} are different in the different microcrystals and average to zero. Consequently, in a polycrystalline medium, \mathbf{P} and \mathbf{E} are parallel. The medium is macroscopically isotropic.

Macroscopically anisotropic media are not only represented by macroscopic crystals, but also glasses or plastic materials if they are produced in a soft state and then hardened under stress in a privileged direction, as is often the case.

4.7 Electronic and Orientation Polarization

In this section, we shall try to understand the dielectric polarization phenomena from the microscopic point of view, extending the hints already given in Sect. 4.2. We immediately state that the correct description of the atomic phenomena is a quantum description. The classic arguments that we shall expose can, however, be considered a reasonably correct first approximation.

There are two main classes of dielectric: the polar and the nonpolar. The nonpolar dielectrics are those we have discussed so far. Their molecules do not have a permanent dipole moment but develop a moment as a reaction to an electric field. The molecules of the polar dielectrics do have a permanent dipole moment. These moments have completely chaotic directions in absence of an applied field and get partially oriented in the presence of a field. Let us start with the first class.

Nonpolar dielectrics

As we have seen, the dipole moment developed by a molecule in the presence of an external field is given by

$$\mathbf{p} = \alpha \mathbf{E}_a. \quad (4.27)$$

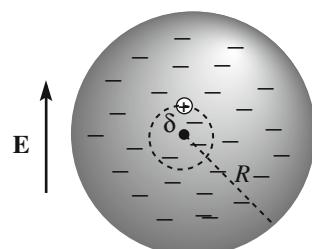
Compared with Eq. (4.8), we have now been more precise. The field inducing the dipole moment in the molecule, which we call an acting field and indicate with \mathbf{E}_a , is conceptually different from the electric field \mathbf{E} that we apply. Let us fix our attention on one particular molecule. It is surrounded by an enormous number of sister molecules. When we apply the field, each of them becomes a dipole and each dipole produces its own small electric field. The fields of all these dipoles, small but very numerous, act upon our molecule together with the applied field. The difference between an externally applied and an acting field is important in condensed media. For the moment, we limit the discussion to the simpler case of a gas, in which the distances between the molecules are large enough that we can neglect their contributions to the acting field. As a further simplification, we shall consider a monoatomic gas. We shall come back to condensed media in the next section.

In the absence of an external field, the atoms of the medium can be considered as a spherical “cloud” of negative charge $-Zq_e$ ($-q_e$ is the electron charge and Z is their number) with an equal and opposite point charge $+Zq_e$ in the center (the nucleus). To make things simple, we assume the negative charge density to be uniform inside the atom. The dipole moment is zero. As a reaction to an external field, the centers of the negative and positive charges displace one another by a distance δ , as shown in Fig. 4.11.

Let us assume the effect of the electric field \mathbf{E} to be simply a rigid shift by δ of the positive charge. The nucleus is no longer in the center of the negative charge. Let Q_{int} be the (negative) charge in the central negative sphere of radius δ , and \mathbf{F} the force it exerts on the nucleus. The direction of \mathbf{F} is opposite to that of the acting field. The magnitude of \mathbf{F} is calculated considering Q_{int} concentrated in the center, obtaining

$$F = \frac{Zq_e Q_{\text{int}}}{4\pi\epsilon_0 \delta^2}.$$

Fig. 4.11 Cartoon representing an atom in an electric field. The dot is the center of the negative charges



Let us express Q_{int} as a function of δ . If R is the radius of the atom, we have the proportion $Q_{\text{int}}/(Zq_e) = (4\pi\delta^3/3)/(4\pi R^3/3)$, and hence, $Q_{\text{int}} = Zq_e\delta^3/R^3$. The equilibrium is reached when δ is such that F is equal to the force due to the acting field Zq_eE_a , namely

$$Zq_eE_a = \frac{(Zq_e)^2\delta}{4\pi\epsilon_0 R^3},$$

from which we immediately have

$$E_a = \frac{Zq_e\delta}{4\pi\epsilon_0 R^3} = \frac{p}{4\pi\epsilon_0 R^3}. \quad (4.28)$$

where p is the induced dipole moment, namely the distance between the centers of the charges times the charge Zq_e .

As we expected, the dipole moment is proportional to the electric field and has its direction and sense. As we noticed in Sect. 4.2, the ratio a/ϵ_0 has the dimensions of a volume. The correlated quantity

$$\alpha_V = \frac{\alpha}{4\pi\epsilon_0} \quad (4.29)$$

is considered in its place for historical reasons. This is called the polarizability volume. Our very rough model provides the estimate

$$\alpha_V = R^3.$$

The atomic radiiuses being on the order of 0.1 nm, α_V should be on the order of 10^{-30} m^3 . Table 4.1 reports the α_V for a number of atoms.

The macroscopic quantity related to α_V is the electric susceptibility

$$\chi_e = \kappa - 1 = n_p 4\pi \alpha_V.$$

Let us recall that the molecular number density at STP of a gas is $n_p = 2.69 \times 10^{25} \text{ m}^{-3}$. For He gas, we calculate $\kappa - 1 = 2.69 \times 10^{25} \times 0.21 \times 10^{-30} \times 4\pi \cong 7 \times 10^{-5}$. This is small compared to unity, providing an a posteriori justification for having considered the acting and applied fields equal. The experimental value is $\kappa - 1 = 6.8 \times 10^{-5}$ (see Table 4.4).

Table 4.1 Polarizability volumes of several elements in 10^{-30} m^3

H	He	Li	N	Ne	Na	Al	Ar	K	Cs
0.68	0.21	24.6	1.14	0.4	24.2	6.9	1.65	43.8	60.2

Adapted from Peter Schwerdtfeger; Massey University Aukland <http://ctcp.massey.ac.nz/Tablepol2015.pdf>

As another exercise, let us evaluate the deformation of a helium atom, relative to its radius, for an intense applied electric field, for example, $E_a = 10^6 \text{ V/m}$. From Eq. (4.28) with $Z = 2$, with a typical atomic radius $R = 0.1 \text{ nm}$, we have $\frac{\delta}{R} = \frac{4\pi\epsilon_0 R^2 E_a}{2q_e} = \frac{(10^{-10})^2 \times 10^6}{2 \times 9 \times 10^9 \times 1.6 \times 10^{-19}} = 3.5 \times 10^{-6}$. The deformation is a very small fraction of the radius indeed. Atoms are really very stiff. This is because even a field as considerably intense as 10^6 V/m is still very small when compared to the fields inside atoms that are in the range of 10^{11} V/m . As a matter of fact, we cannot produce macroscopic fields on the order of the atomic ones, simply because, to produce a field, we need electrodes, and electrodes are made of atoms, and the latter would be destroyed if subject to an external field similar to their internal ones. Practical limits are, however, reached at much smaller values. We can conclude that the linearity hypothesis is almost always verified.

Polar dielectrics

Very often, even in some of the simplest molecules, the centers of the positive and negative charges are different. The molecule then has an intrinsic dipole moment, which we shall indicate with \mathbf{p}_0 . An example is the water molecule, which, as we saw in Sect. 1.15, has a large permanent dipole moment. We shall now consider a fluid of polar molecules. In the absence of an applied electric field, the orientations of the molecular dipoles are chaotic with the consequence that, in any volume, even if macroscopically infinitesimal but still containing an enormous number of molecules, the net dipole moment is zero. The effects of an applied electric field are twofold. The first effect is exactly the same as for the nonpolar dielectrics; the applied field induces a deformation and, consequently, a dipole moment, now in addition to the permanent one. The second effect, which is quantitatively larger, is to produce a statistical preference of the permanent dipoles toward directions close or equal to that of the field. This can happen freely in a fluid where the molecules can slide on one another, but not in a solid.

The energy of a dipole \mathbf{p}_0 in the field \mathbf{E}_a , as given by Eq. (1.95), is

$$U_E = -\mathbf{p}_0 \cdot \mathbf{E}_a = -p_0 E_a \cos \theta \quad (4.30)$$

where θ is the angle of the dipole moment with the field. The statistical equilibrium condition is given by the Boltzmann factor (see Vol. 2, Chap. 5), named after Ludwig Boltzmann (Austria, 1844–1906). Namely, the number of molecules per unit volume and per unit solid angle with energy U_E is given by

$$n(\theta) = n_0 e^{-\frac{U_E}{k_B T}} = n_0 e^{+\frac{p_0 E_a \cos \theta}{k_B T}} \quad (4.31)$$

where k_B is the Boltzmann constant, T is the absolute temperature of the fluid and n_0 is a normalization constant that we shall determine. We now observe that, for temperatures not too close to the absolute zero, the exponent in the last equation is

very small. Consequently, we can expand in series and stop at the first term, obtaining

$$n(\theta) = n_0 \left(1 + \frac{p_0 E_a \cos \theta}{k_B T} \right) \quad (4.32)$$

We now determine the normalization constant by imposing that the number of molecules per unit volume having a dipole moment in the entire solid angle must be of a numerical density n_p . This number is

$$\int_0^{4\pi} n(\theta) d\Omega = n_0 \int_0^{2\pi} d\phi \int_0^{\pi} \left(1 + \frac{p_0 E_a \cos \theta}{k_B T} \right) \sin \theta d\theta = 4\pi n_0,$$

which must be equal to n_p . Hence, the normalization constant is $n_0 = n_p / 4\pi$.

The next step is calculating the dipole moment per unit volume \mathbf{P} . Knowing that \mathbf{P} has the direction of the field, we can only consider the components of the molecular moments in that direction. Each of the $d\mathbf{n}$ molecules in the solid angle $d\Omega$ around θ gives a contribution $p_0 \cos \theta$, and all the $d\mathbf{n}$ together a contribution $p_0 \cos \theta d\mathbf{n} = p_0 \cos \theta n(\theta) d\Omega$. Integrating all the solid angles, we have

$$\begin{aligned} P &= \int_0^{4\pi} p_0 n(\theta) \cos \theta d\Omega = \frac{n_p p_0}{4\pi} \int_0^{2\pi} d\phi \int_0^{\pi} \left(1 + \frac{p_0 E_a \cos \theta}{k_B T} \right) \cos \theta \sin \theta d\theta \\ &= -\frac{n_p p_0}{2} \int_{+1}^{-1} \left(1 + \frac{p_0 E_a \cos \theta}{k_B T} \right) \cos \theta d(\cos \theta). \end{aligned}$$

The last integral is immediately calculated, providing

$$P = \frac{n_p p_0^2 E_a}{3 k_B T} \quad (4.33)$$

The measurable quantity is the dielectric constant, for which we obtain the expression

$$\kappa - 1 = \frac{P}{\epsilon_0 E_a} = \frac{n_p p_0^2}{3 \epsilon_0 k_B T} \quad (4.34)$$

This law is well verified by experiment. Notice that the dielectric constant of the polar dielectrics depends on the temperature, as opposed to the nonpolar ones.

Another difference appears under dynamic conditions. As a matter of fact, the main arguments of this section continue to hold even if the field varies in time. However, the time necessary for the deformation of an atomic cloud is much shorter

Table 4.2 Static dielectric constants κ of several solids and plastics at $T = 25^\circ\text{C}$

Si	NaCl	CsCl	PVC	Neoprene	Polyethylene	Teflon AF
11.9	5.90	7.20	7.0	9.8	2.3–2.7	2.1

Table 4.3 Static dielectric constants κ of several liquids at $T = 25^\circ\text{C}$

BR ₂	C ₆ H ₆	CCl ₄	CH ₄ O	H ₂ O	NH ₃	C ₆ H ₆
3.09	2.27	2.23	32.63	78.54	16.9	2.27

Table 4.4 Static dielectric constants ($\kappa - 1$) in 10^{-4} units of several gases at atmospheric pressure (101.3 kPa) and $T = 20^\circ\text{C}$

He	Ne	Ar	H ₂	N ₂	O ₂	Air	CO ₂	N ₂ O
0.68	1.3	5.16	2.54	5.47	4.94	5.36	9.88	11.0

than the typical time needed to reach the statistical equilibrium of the dipole orientations in a polar dielectric. If we apply a periodically varying field, when its frequency is low, the dipoles have time enough to orient themselves following the field and the polar dielectric behaves almost as if under static conditions. However, if the frequency increases above a certain value, the dipoles cannot reorient any more. Above these frequencies, which are typically on the order of the GHz, only the deformation effect remains. If we measure the dielectric constant as a function of frequency of the applied field, we observe a transition from values close to those under a static condition to much lower values typical of a nonpolar dielectric.

Tables 4.2, 4.3 and 4.4 report several examples of dielectric constants under static conditions for solids and plastics, liquids and gases, respectively. Notice that, at STP, the dielectric constants of gases differ from one another only by a few parts in 10,000. The reported values are for $\kappa - 1$ in 10^{-4} units.

4.8 Electric Field in Cavities in Dielectrics

We shall now consider the problem of the dielectric constant in a dense medium. As already mentioned, the problem is complex, because the polarization of each molecule is a consequence not only of the applied field, but also of the fields of the polarizations of all the other molecules. To evaluate this *acting field*, let us consider one of the molecules, namely the molecule on which we want to find \mathbf{E}_a . Let us imagine taking the molecule out of the medium, and doing that without changing the state of the medium, leaving, so to speak, its polarization frozen. The acting field is the field in the cavity we have opened. The problem is not simple, because

the field in a cavity in a dielectric depends on the shape of the cavity. Let us analyze this issue, which is interesting from other points of view as well.

Let us start with a narrow cylindrical cavity with axis parallel to \mathbf{P} and \mathbf{E} . We take its height to be much longer than the diameter of its base, as in Fig. 4.12. The question is: if \mathbf{E} is the field in the dielectric, what is the field \mathbf{E}_0 in the cavity? Now, \mathbf{E}_0 is different from \mathbf{E} if, in opening the cavity, we have altered the divergence or the curl of the field. We did not alter the curl, which under static conditions is always zero. A few polarization charges have appeared on the faces of the cylinder with a possible effect on the divergence. However, these charges are very few, because we made the faces very small, and they are far from the center of the cylinder, because we made it long. In conclusion, the electric fields in the cavity and in the dielectric are equal, namely

$$\mathbf{E}_0 = \mathbf{E} \quad (4.35)$$

Consider now a flat cylindrical cavity with a height much shorter than the diameter of the base, again with the axis parallel to \mathbf{P} and \mathbf{E} , as in Fig. 4.13. The contribution of the polarization charges appearing on the faces are now important, because the faces are large and near to the center. If σ_P is the polarization charge density, its contribution to the electric field is σ_P/ϵ_0 or, better yet, in vector form, \mathbf{P}/ϵ_0 , which has the same direction of \mathbf{E} .

The field in the cavity is

$$\mathbf{E}_0 = \mathbf{E} + \mathbf{P}/\epsilon_0. \quad (4.36)$$

Fig. 4.12 The field in a narrow cavity in a dielectric

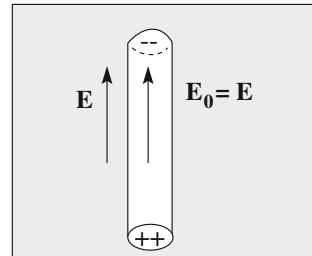
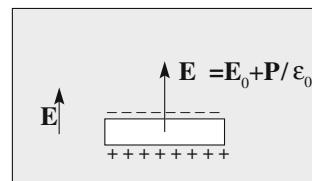


Fig. 4.13 The fields in a flat cavity in a dielectric



We can reach the same result comparing the field \mathbf{D} in the dielectric with the same field, which we call \mathbf{D}_0 , in the cavity. As we know, the divergence of \mathbf{D} is due only to the free charge, and consequently

$$\mathbf{D}_0 = \mathbf{D}, \quad (4.37)$$

which is equivalent to Eq. (4.36).

We can now go back to the field acting on a molecule. We make the approximation that the acting field, at least for liquid dielectrics with molecules that are not too structurally complex, is the field in the center of a spherical cavity.

Consider a dielectric in a uniform field. Let us start by ideally cutting a sphere in the dielectric, but not yet removing it. The field \mathbf{E} in the sphere is obviously equal to the field at any other point of the dielectric. We use the superposition principle, as in Fig. 4.14, by saying that \mathbf{E} in the sphere is equal to the sum of the field of the polarized sphere taken alone (call it \mathbf{E}_S) and the field of the remaining dielectric with a spherical hole in its center. The latter is the field \mathbf{E}_a that we are searching for. We have then $\mathbf{E}_a = \mathbf{E} - \mathbf{E}_S$.

We will postpone to the end of the section the calculation of the field of a polarized sphere, so as not to interrupt the discussion. The result is

$$\mathbf{E}_S = -\frac{\mathbf{P}}{3\epsilon_0}. \quad (4.38)$$

Hence, we have

$$\mathbf{E}_S = \mathbf{E} + \frac{\mathbf{P}}{3\epsilon_0}, \quad (4.39)$$

As we see, this is an intermediate value between the long and flat cavities. Considering that $\mathbf{P} = n_p \alpha \epsilon_0 \mathbf{E}_a$ and using the just found \mathbf{E}_S for \mathbf{E}_a , we get $\mathbf{P} = n_p \alpha \epsilon_0 (\mathbf{E} + \mathbf{P}/3\epsilon_0)$, namely

$$\mathbf{P} = \frac{n_p \alpha}{1 - n_p \alpha / 3} \epsilon_0 \mathbf{E}. \quad (4.40)$$

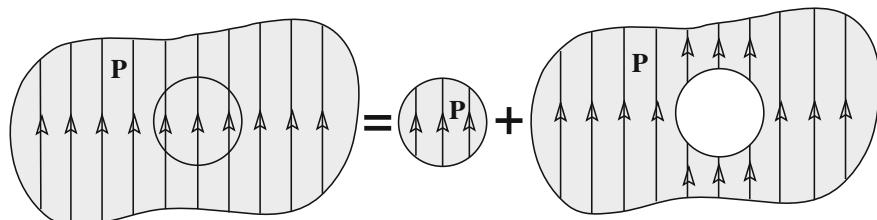


Fig. 4.14 Combining the contributions to the field in a spherical cavity

Finally, for the dielectric constant, we get

$$\kappa = 1 + \chi_e = 1 + \frac{n_p \alpha}{1 - n_p \alpha / 3}. \quad (4.41)$$

This equation expresses a macroscopic quantity, the dielectric constant in terms of atomic quantities. It was explicitly written for the first time by Rudolf Clausius (Germany, 1822–1888) in 1879 after having been found, in a different form, by Ottaviano Mossotti (Italy, 1791–1863) back in 1850, and is thus known as the *Clausius-Mossotti formula*.

Notice that, for low-density media, such as a gas at STP, namely for small values of n_p , Eq. (4.41) gives back $\chi_e = n_p \alpha$, which we already met. On the other hand, Eq. (4.41) shows that, for dense media, namely for large values of n_p , the dielectric constant can be significantly larger than the unit.

Let us now compare Eq. (4.41) with a few experimental data. Consider, for example, CS_2 . In the gas phase at STP, the dielectric constant is measured to be $\kappa = 1.0029$. With this value, we calculate from Eq. (4.41) $n_p \alpha = 0.0029$. Consider now the liquid phase. α does not vary, but n_p is larger, like the ratio between liquid and gas densities, which, for CS_2 , is 381. We then have $n_{p,\text{liq}} \alpha = 0.0029 \times 381 = 1.1$. With this value, Eq. (4.41) predicts the value of the dielectric constant to be $\kappa = 2.75$ (which is significantly larger than 1), while the measured value is $\kappa = 2.64$. The agreement is even better than that which one might expect from our very rough model. In general, the Clausius-Mossotti formula gives values within 10–20 % of the experimental values, as long as nonpolar dielectrics are considered. Contrastingly, its predictions are completely wrong for polar dielectrics, as expected, because the underlying model completely ignores the presence of molecular dipoles.

Let us now calculate the electric field of a uniformly polarized dielectric sphere. We choose the origin of the reference frame to be in the center of the sphere and the z -axis in the positive direction of the polarization \mathbf{P} . Let θ be the polar angle and \mathbf{n} the outgoing unit vector normal to the surface of the sphere. Notice that \mathbf{n} has the direction of the radius and that θ is also the angle between \mathbf{n} and \mathbf{P} . Since the polarization is uniform, the only polarization charges are on the surface, as shown in Fig. 4.15, with density

$$\sigma_p = \mathbf{P} \cdot \mathbf{n} = P \cos \theta. \quad (4.42)$$

Let us now consider having two spheres equal to the one being considered, charged with uniform volume charges of opposite sign. Let Q be the positive charge. Let the two spheres initially overlap perfectly and let us shift the positive sphere in the z direction by a small distance δ , as shown in Fig. 4.16. We obtain two crescents in which the charges no longer overlap and which are both charged, one positively and one negatively. The charge under the unit surface (namely the surface charge density) is proportional to the thickness of the crescent at that point, which is, in turn, proportional to $\cos \theta$. We conclude that the system of our two

Fig. 4.15 The electric field of a uniformly polarized sphere

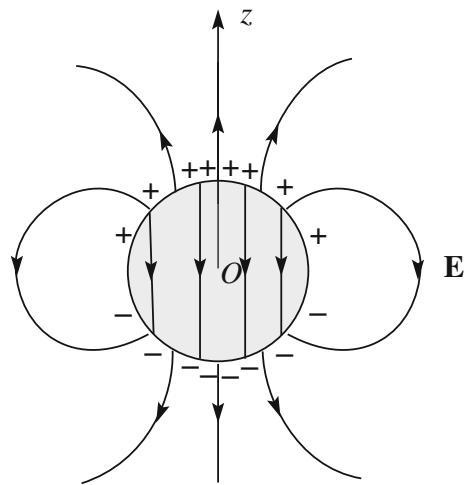
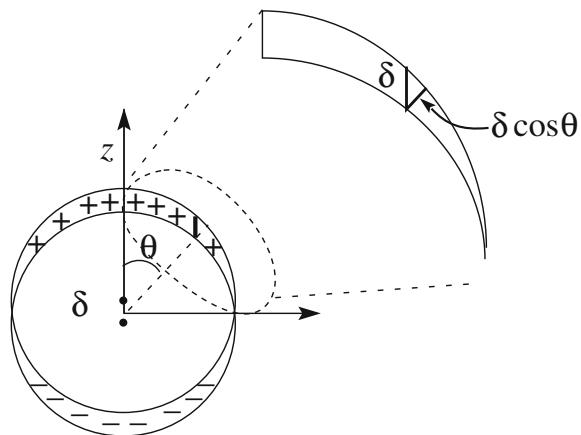


Fig. 4.16 Two oppositely charged, slightly displaced spheres



spheres is equivalent to the surface density of Eq. (4.42), provided we properly chose Q and δ .

The field of the two spheres at the external points is equal to the field of two point charges $+Q$ and $-Q$ in the centers of the spheres. This last system is a dipole with a dipole moment of magnitude $Q\delta$ and the direction of z -axis.

To be equivalent to polarized sphere, this moment should be equal to the dipole moment of the polarized sphere, namely $Q\delta = (4/3)\pi R^3 P$. In conclusion, the electric field of the polarized sphere outside the sphere is equal to the field of an electric dipole at the origin having the moment $\mathbf{p} = (4/3)\pi R^3 \mathbf{P}$. It is convenient here to use the potential, in place of the field, and polar coordinates (r, θ) . The potential outside the sphere is

$$\phi(r, \theta) = \frac{1}{4\pi\epsilon_0} \frac{p \cos \theta}{r^2} = \frac{P}{3\epsilon_0} \frac{R^3}{r^2} \cos \theta. \quad (4.43)$$

In particular, the surface of the sphere (recalling that $R \cos \theta = z$) is

$$\phi(r, \theta) = \frac{P}{3\epsilon_0} R \cos \theta = \frac{P}{3\epsilon_0} z. \quad (4.44)$$

We see that it depends only on z , namely the equipotentials on the surface are the geographic parallels on the sphere. From this consideration, we can immediately obtain the potential inside the sphere. Indeed, as no charge is present in the volume, the potential must be the solution to the Laplace equation with the boundary conditions on the surface given by Eq. (4.44). But, as immediately verified, Eq. (4.44) is indeed a solution to the Laplace equation inside the sphere with those boundary conditions. The solution being unique, this is also *the* searched-for solution inside the sphere. The electric field only has a non-zero component, namely the z -component, which is

$$E_z = -\frac{P}{3\epsilon_0}, \quad (4.45)$$

which is Eq. (4.38); in particular, \mathbf{P} is uniform inside the sphere.

4.9 Electrostatic Energy in a Dielectric

In Chap. 3, we discussed the energy of systems of charges in fixed positions in a vacuum and of systems of conductors also in a vacuum. We now discuss the energy when the space between conductors is filled, at least partially, with a dielectric, which we shall assume to be normal.

Consider a system of conductors in given fixed positions, each with a certain charge. As is always the case, the energy of the system is equal to the work we must spend to build it. Let us start with no charges on the conductors. We charge each conductor in turn by bringing in charges from infinity to its surface until they have reached the desired values. We shall call these free charges, as usual. To bring in the charges, we must provide work against the field of the free charges that are already in place, exactly as if there were a vacuum. However, as opposed to when there was a vacuum, now the dielectric polarizes as we bring charges in. In the process, the molecules of the medium deform and reorient in a process taking additional external work.

Let us start by considering a parallel plate capacitor with plates of surface S and distance d . We start with a vacuum between the plates. We connect the plates to a battery, charging them with $+Q$ and $-Q$, respectively, then we disconnect the battery. The work done on the system, which is the electric work done by the battery, is

$$U_E^0 = \frac{1}{2} \frac{Q^2}{C_0} = \frac{1}{2} Q^2 \frac{h}{\epsilon_0 S} \quad (4.46)$$

where we used the superscript and subscript 0 to indicate that we are in a vacuum.

Let us now have a normal dielectric of dielectric constant κ between the plates.

The capacitance is now larger by a factor κ . As a consequence, to charge the capacitor with the same Q as before, we shall connect it to a battery of a lower potential difference. The work done on the system is

$$U_E = \frac{1}{2} \frac{Q^2}{C} = \frac{1}{2} Q^2 \frac{h}{\kappa \epsilon_0 S}. \quad (4.47)$$

Being that $U_E < U_E^0$, the extra work done to polarize the dielectric is negative. Let us understand the reason for that by thinking of building the system in a different way. Suppose the dielectric to be a plastic plate. We start by charging the capacitor in a vacuum. The work done by the battery is given by Eq. (4.46). We disconnect the battery, to work at constant free charge, and then we introduce the dielectric plate. This requires mechanical work. Here, the edge effects, which we have neglected in other instances, are crucial. Figure 4.17 shows the electric field lines when the dielectric plate is partially inserted. Notice that the field lines are different on the right-hand side, where they are in a vacuum, than on the left-hand side, where they penetrate the dielectric. Notice also that the lines change direction when crossing the dielectric surface, due to the presence of polarization surface charges.

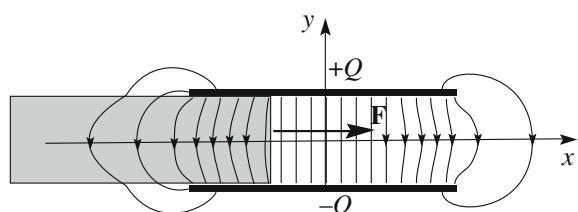
Entering into the electric field, the dielectric becomes polarized, namely every infinitesimal volume dV becomes a dipole of moment $\mathbf{P}dV$, where \mathbf{P} is the polarization density (which is a function of the point). The force exerted by an electric field on a dipole is non-zero only if the field is not uniform. But this is just the case in the fringe field of the capacitor. The force on the dipole in dV is given by Eq. (1.97), namely

$$d\mathbf{F} = (\mathbf{P}dV \cdot \nabla)\mathbf{E}, \quad (4.48)$$

which we can write, using Eq. (4.11), as

$$d\mathbf{F} = \epsilon_0 \chi_e (\mathbf{E} \cdot \nabla) \mathbf{E} dV. \quad (4.49)$$

Fig. 4.17 Force on a dielectric plate being introduced in a capacitor



We now use the vector identity $\nabla \mathbf{E}^2 = 2(\mathbf{E} \cdot \nabla) \mathbf{E}$ and write Eq. (4.49) as

$$d\mathbf{F} = \frac{1}{2} \epsilon_0 \chi_e \nabla \mathbf{E}^2 dV. \quad (4.50)$$

The force on the dielectric is obtained by integration on its volume V as

$$\mathbf{F} = \frac{1}{2} \epsilon_0 \chi_e \int_V \nabla \mathbf{E}^2 dV. \quad (4.51)$$

Looking at Fig. (4.17), we see that the force component F_z is zero, because E_z is identically zero, and that the component F_y is zero as well, because the contributions of the volume elements located symmetrically relative to the median plane cancel each other out. Hence, the only non-zero component is

$$F_x = \frac{1}{2} \epsilon_0 \chi_e \int_V \frac{\partial \mathbf{E}^2}{\partial x} dV. \quad (4.52)$$

Unfortunately, we cannot calculate further, because we do not know $\mathbf{E}(x, y)$, which is a quite complicated function. We can, however, see that $F_x > 0$, namely that the force is directed inside (similarly to an electrostatic pendulum). Indeed, $\partial \mathbf{E}^2 / \partial x > 0$ everywhere in the region of the non-uniform field and, of course, $\partial \mathbf{E}^2 / \partial x = 0$ well inside where the electric field is uniform. Consequently, $F_x > 0$. The integral Eq. (4.52), namely the work to be done against this force to bring the dielectric into position, is negative.

Let us now express energy as energy of the fields. We continue with a parallel plate capacitor, letting σ_f be the free charge surface density, S the area of the plates and h their distance. Neglecting the edge effect, the field is uniform in the volume $V = Sh$ and zero outside. We can write Eq. (4.47) in the equivalent form as

$$U_E = \frac{1}{2} \frac{Q^2}{C} = \frac{1}{2} \frac{\sigma_f \sigma_f}{\epsilon_0 \kappa} Sh = \frac{1}{2} \mathbf{E} \cdot \mathbf{D} V.$$

We interpret this expression by saying that there is energy in the field with density

$$w_E = \frac{1}{2} \mathbf{D} \cdot \mathbf{E}. \quad (4.53)$$

The field energy is then

$$U_E = \frac{1}{2} \int \mathbf{D} \cdot \mathbf{E} dV, \quad (4.54)$$

which takes the place of

$$U_E = \frac{\epsilon_0}{2} \int \mathbf{E}^2 dV. \quad (4.55)$$

Notice that in a vacuum, $\mathbf{D} = \epsilon_0 \mathbf{E}$ and the two expressions coincide.

We now show that the Eqs. (4.53) and (4.54) that we have found in a particular case hold for any linear dielectric. For simplicity's sake, we shall assume that it is also isotropic. Then, the relation between \mathbf{D} and \mathbf{E} is simply $\mathbf{D} = \epsilon \mathbf{E}$.

Let us consider a region of space in which a dielectric and a number of free charges are present. To calculate the energy, we start from the state in which there are no free charges on the conductors and evaluate the work necessary to bring the charges from the infinite. Let $\rho_f(\mathbf{r})$ be the free charge density at the point \mathbf{r} in the generic step of the charging operation. Under these conditions, the dielectric is already (partially) polarized, hence we have to deal with polarization charges. Let $\phi(\mathbf{r})$ then be the potential due to all the charges, both the free and the polarization ones. Let us now vary the free charge density already in place by $\delta\rho_f(\mathbf{r})$, by bringing the charge $\delta\rho_f(\mathbf{r})dV$ from infinite in the elementary volume dV . The work to be done by the external agent on the system is

$$\delta W = \int_{\text{all space}} \delta\rho_f(\mathbf{r})\phi(\mathbf{r})dV,$$

where we have extended the integral to the entire space, not only to the volume of the charge distribution, because the integral is zero outside that volume and because this form will be immediately useful. Recall now that $\nabla \cdot \mathbf{D} = \rho_f$. Hence, if ρ_f varies by $\delta\rho_f$, then \mathbf{D} varies by $\delta\mathbf{D}$ such that $\nabla \cdot \delta\mathbf{D} = \delta\rho_f$. Substituting this into the above equation, we get

$$\delta W = \int_{\text{all space}} \phi \nabla \cdot \mathbf{D} dV.$$

We now use the identity $\phi \nabla \cdot \mathbf{D} = -\nabla\phi \cdot \delta\mathbf{D} + \nabla \cdot (\phi \delta\mathbf{D})$ and recall that $\mathbf{E} = -\nabla\phi$, obtaining

$$\delta W = \int_{\text{all space}} \mathbf{E} \cdot \delta\mathbf{D} dV + \int_{\text{all space}} \nabla \cdot (\phi \nabla \cdot \mathbf{D}) dV.$$

Exactly as we did in Sect. 3.5 in the case of a vacuum, we transform the second integral on the right-hand side into a surface integral with the Gauss divergence theorem and recognize that it vanishes when the surface goes to infinity. In this way, we are left with

$$\delta W = \int_{\text{all space}} \mathbf{E} \cdot \delta \mathbf{D} dV. \quad (4.56)$$

We obtain the energy of the system, namely the total work needed to build it, by integrating this expression for \mathbf{D} , changing it from zero to its final value. Up to now, we have not made any hypothesis on the dielectric. We now need a relation between \mathbf{D} and \mathbf{E} and need to assume it to be $\mathbf{D} = \epsilon \mathbf{E}$, namely to deal with a normal dielectric. By differentiation, we then obtain $\delta(\mathbf{D} \cdot \mathbf{E}) = \epsilon \delta(\mathbf{E} \cdot \mathbf{E}) = \epsilon \mathbf{E} \cdot \delta \mathbf{D}$. In the last term, we have the integrand of Eq. (4.56). We can then write for the total energy

$$U_E = \int_0^D \delta W = \frac{1}{2} \int_{\text{all space}} dV \int_0^D \delta(\mathbf{D} \cdot \mathbf{E}) = \frac{1}{2} \int_{\text{all space}} \mathbf{D} \cdot \mathbf{E} dV,$$

which demonstrates the thesis.

Let us recall that, in Sect. 3.5, we had found Eq. (4.55) starting from the equation

$$U_E = \frac{1}{2} \int \rho \phi dV. \quad (4.57)$$

With the same argument from Sect. 3.5, only reversed, it is shown that the energy of the free charges immersed in a normal dielectric can be expressed as

$$U_E = \frac{1}{2} \int \rho_f \phi dV. \quad (4.58)$$

Equations (4.54) and (4.58) are equivalent in electrostatics. In the former, the energy appears to be distributed in space, and in the latter, to be localized on the free charges. In practice, depending on the problem, one or the other equation may be more useful.

Equation (4.54) is the total work done to assemble the system. The external agent acts directly on the free charges alone. However, the force the agent needs to exert on the free charges depends on the arrangement of the polarization charges, which changes during the assembly process. Albeit indirectly, part of the work goes into stretching and re-orienting the molecules. As we have already noticed, Eq. (4.58) looks very similar to Eq. (4.57), but the similarity is misleading. In Eq. (4.57), in a vacuum, ϕ is the potential of the free charges in their own field. The 1/2 factor is there to correct for the double counting. Contrastingly, in Eq. (4.58), ϕ is the potential of all the charges, both the free and the polarization ones. In addition, the origin of the 1/2 factor is completely different. Let ϕ_f and ϕ_P be the potentials generated by the free and polarization charges, respectively, and let us split the two contributions in Eq. (4.58), writing

$$U_E = \frac{1}{2} \int \rho_f \phi_f dV + \frac{1}{2} \int \rho_f \phi_P dV = U_f + U_X. \quad (4.59)$$

The first term on the right-hand side is the energy of the free charges in their own field. The factor 1/2 is to cancel the double counting. However, the second term, namely what we called U_X , is 1/2 of the energy of the free charges in the field of the polarization charges. Why one half only? In addition, in Eq. (4.59), the energy of the polarization charges in their own field does not explicitly appear. To try to make a complicated situation clear, let us start with bookkeeping. The energies to be considered are the following.

The energy of the free charges in their own field is, as we already noticed,

$$U_f = \frac{1}{2} \int \rho_f \phi_f dV. \quad (4.60)$$

Similarly, the energy of the polarization charges in their own field is

$$U_P = \frac{1}{2} \int \rho_P \phi_P dV. \quad (4.61)$$

The third and last term is the interaction energy between free and polarization charges. This can be written in two equivalent forms. In one form, the free charges are seen as the sources of the field and the polarization charges as the receptors of the field force; in the other way, the roles are inverted, but the energy is the same. Namely the term is

$$U_{\text{int}} = \int \rho_f \phi_P dV = \int \rho_P \phi_f dV, \quad (4.62)$$

which is twice U_X in Eq. (4.59).

Let us now build the system with a different procedure, in two steps, as we already did in the particular case of the parallel plate capacitor. The first step is to charge the conductors in a vacuum, bringing the (free) charges from infinity on them. The work needed for that, and the corresponding energy, is U_f . The second step is to introduce the dielectric medium in position, bringing it in from infinity. The charges on the conductors are kept fixed and no work is done on them. To calculate the work needed to bring in the dielectric, we must consider the electric field before and after this process. Let \mathbf{E}_0 be the electric field in a vacuum, namely without the dielectric, and \mathbf{E} and \mathbf{D} the electric and electric displacement fields in the final state. The work we are considering is equal to the difference between the energy of the system before and after having brought the dielectric in, namely

$$W = \frac{1}{2} \int (\mathbf{E} \cdot \mathbf{D} - \epsilon_0 \mathbf{E}_0^2) dV.$$

We now want to have \mathbf{E}_0 only on the right-hand side. From Eq. (4.54), we have $\mathbf{E} = \mathbf{E}_0/\kappa$, and from Eqs. (4.11) and (4.13), $\mathbf{P} = \epsilon_0(\kappa - 1)\mathbf{E} = \epsilon_0 \frac{\kappa-1}{\kappa} \mathbf{E}_0$. Substituting in the above expression, we get

$$W = -\frac{1}{2} \int \mathbf{E}_0 \cdot \left(\epsilon_0 \frac{\kappa-1}{\kappa} \right) \mathbf{E}_0 dV = -\frac{1}{2} \int \mathbf{P} \cdot \mathbf{E}_0 dV = \frac{1}{2} \int \mathbf{P} \cdot \nabla \phi_f dV$$

We now use the vector identity $\nabla \cdot (\phi_f \mathbf{P}) = \mathbf{P} \cdot \nabla \phi_f + \phi_f \nabla \cdot \mathbf{P}$, obtaining

$$W = \frac{1}{2} \int_{\text{all space}} \nabla \cdot (\phi_f \mathbf{P}) dV - \frac{1}{2} \int_{\text{all space}} \phi_f \nabla \cdot \mathbf{P} dV$$

Similarly to what we have already done several times, we can express the first term on the right-hand side using the Gauss divergence theorem as a surface integral on the surface surrounding the volume and sending the surface to infinity. The limit is zero. In the second term, we notice that $-\nabla \cdot \mathbf{P} = \rho_P$, and finally obtain

$$W = \frac{1}{2} \int \rho_P \phi_f dV. \quad (4.63)$$

This is the work to be spent to polarize the dielectric. As we have already noticed, the total electrostatic energy U_E in Eq. (4.58) is equal to the total work done on the system from outside, which is on the free charges alone. The work W needed to polarize the dielectric is done by the free charges at the expense of the interaction energy between the free and polarization charges. The interaction energy decreases by the same amount, namely by $-W$, reducing to U_X in Eq. (4.59). This energy equal to W is hidden in the microscopic field, namely in the molecules that have been stretched and/or re-oriented in the polarization process.

We finally warn the reader that, for the non linear dielectrics, the energy of the final configuration must be calculated from Eq. (4.56) knowing the dependence of \mathbf{D} from \mathbf{E} , which is, in general, very complex.

QUESTION Q 4.3. A capacitor is immersed in oil, connected to a battery and, when charged, disconnected. Its energy is 30 μJ . The oil is then extracted, spending work of 80 μJ . Find the dielectric constant of the oil. Where has the work gone? \square

QUESTION Q 4.4. Find the energy density in a fluid normal dielectric of constant $\kappa = 3$ at 1 m distance from: (a) a 1 pC point charge, (b) a uniform straight charge distribution of 1 pC/m, (c) a uniform plane charge distribution of 1 pC/m². \square

Summary

In this chapter, we have studied electrostatics in the presence of dielectric media. We have learned the following principal concepts:

1. The polarization of the medium and its dielectric constant.
2. The relation between the vector polarization density and the electric field.
3. The relation between the surface and volume charge densities on one side and the polarization vector on the other.
4. The electric displacement vector and its sources.
5. The differential equations for the electrostatic field in a dielectric.
6. The Coulomb force between two point charges at rest in a normal fluid dielectric.
7. The microscopic interpretation of electronic and orientation polarization phenomena.
8. The electric fields in a cavity in a dielectric and how they depend on the shape of the cavity.
9. The charges and field energy in a dielectric.

Problems

- 4.1. Fig. 4.1 shows polarization charges on the surfaces of the plates inside the capacitor and a discontinuity of the electric field. Is the \mathbf{D} field continuous or discontinuous?
- 4.2. With reference to Fig. 4.1, what is the charge density inside the dielectric?
- 4.3. Can the electric conductivity be negative? Why?
- 4.4. What is the order of magnitude of the thickness of the surface polarization charge density?
- 4.5. Why is the dielectric constant of water large?
- 4.6. Does the dielectric constant of water depend on temperature? And that of helium?
- 4.7. Consider a cylindrical cavity in a dielectric with axis parallel to the electric field. Is the electric field larger in a long narrow cavity or in a flat wide one? And the electric displacement?
- 4.8. Two plane parallel surfaces in a vacuum are charged with equal and opposite surface densities. We fill the space between them with a normal dielectric of constant κ . How do the electric field \mathbf{E} , the electric displacement \mathbf{D} and the potential difference change?
- 4.9. Consider two plane surfaces at a distance d in a dielectric. Let \mathbf{n} be the unit vector normal to them. The polarization on one of the surfaces is \mathbf{P}_1 in the direction and sense of \mathbf{n} , while on the second, it is \mathbf{P}_2 in the direction and sense of \mathbf{n} as well.
- 4.10. A dielectric slab is immersed in an electric field perpendicular to its faces. Consider a closed surface, partially inside and partially outside the slab. The flux of \mathbf{E} outgoing from this surface is positive, while the flux of \mathbf{D} is zero. What can you tell about the charges on the face crossed by that surface?

- 4.11. Evaluate the dielectric constant of liquid Ar, κ_{liq} , knowing that the dielectric constant of Ar gas at STP is $\kappa_{\text{gas}} = 1.000545$ and that the ratio of the liquid to gas (STP) densities is 810. (Compare with the measured value $\kappa_{\text{liq}} = 1.54$.)
- 4.12. In a point inside a dielectric, the dielectric constant is κ and the electric displacement \mathbf{D} . What is the polarization \mathbf{P} ?
- 4.13. What is the charge density at 1 m distance from a small conductive sphere with charge $Q = 1 \text{ pC}$ immersed in a fluid normal dielectric with dielectric constant $\kappa = 2$? Consider the medium to be infinite.
- 4.14. The closed surface S in a dielectric does not contain free charges. The flux of the polarization vector outgoing S is 5 pC. What can you say about the polarization charge contained in S ?
- 4.15. In a certain dielectric medium, the polarization as a function of position is $\mathbf{P} = x^2\mathbf{i} + 3yx\mathbf{j} - 5xz^3\mathbf{k}$. Find the expression of the polarization charge density.

Chapter 5

Electric Currents

Abstract In this chapter, we study the direct electric currents in conductors. We define the current intensity, which is a scalar quantity, and the current density, which is a vector quantity. We shall learn the universal law of electric charge conservation, which holds in the strong form of local conservation. The concepts of electric resistance, resistivity and electromotive force (emf, for short) will be introduced. We shall describe, as an historical example, how the British Association standard of resistance was developed. The Ohm laws and the Kirchhoff rules will be discussed.

In this chapter we study the electric currents in the conductors. We shall see how electric current is due to the motion of the charge carriers, electrons or ions for example. We shall define the current intensity, which is a scalar quantity, and the current density, which is a vector quantity. The charge carriers inside a conductor are permanently in motion. In the absence of a macroscopic electric field, the motion is chaotic. A conduction current exists if, in addition, an ordered component of the motion exists. As an analogy, think of to the air molecules in a room. When the room is closed, the molecule's motion is chaotic, but, if we open two windows on two different walls, a draft may develop. The motion of the molecules have now has an ordered component.

In Sect. 5.2, we learn that the electric charge is always conserved. This is a very important, and universal, law of physics. In addition, charge conservation is a local property. This means that charges can vary in a given volume only as a result of charges entering into or coming out of the volume. Charge conservation should not be confused with charge invariance, namely with the property of a charge having the same value in every reference frame, which we studied in Sect. 1.9.

In Sect. 5.3, we introduce the concept of electric resistance, which measures how strongly, so to speak, a conductor opposes the passage of current. As a matter of fact, resistance is due to dissipative phenomena that develop the motion of the carriers tending toward disorder. To maintain the current at a steady value, a potential difference must be applied to the ends of the conductor. The law governing the phenomenon is Ohm's law.

Section 5.4 is a historical interlude that briefly describes how a prototype resistor and the connected measuring procedures were developed in the mid-19th century and their practical and purely scientific importance.

In Sect. 5.5, we deal with Ohm's law for a current running on a surface, rather than in a volume. In Sect. 5.6, we study the energy balance for the passage of a current in a circuit. In the subsequent section, we consider the generators of continuous electromotive force (or simply of voltage) and see how they act on the carriers with non-electrostatic forces.

In Sect. 5.8, we analyze the example of a circuit containing a capacitor, a resistor and a generator, and the processes of charging and discharging the capacitor. In Sect. 5.9, we give the rules for calculating the potential differences and the current intensities in resistive networks in a steady regime (the Kirchhoff rules).

We shall close the chapter with a few hints on superconductivity, which is the property of certain materials having zero resistance at temperatures below one well-defined, called the transition temperature of the material. This is a purely quantum phenomenon, of which we can only give a few ideas here.

5.1 Current Intensity and Current Density

Up to now, we have studied electric phenomena under static conditions, namely ones in which electric charges are at rest and the electric field is constant in time. Under these conditions, in particular, the electric field and the charge density are zero inside homogeneous conductors. We shall now study phenomena in which electric charges move over macroscopic distances, producing an *electric current*.

Let us start by considering two metallic conductors initially isolated and without charge. We move a negative charge from conductor 2 to conductor 1. Let ϕ_1 and ϕ_2 with $\phi_1 < \phi_2$ be the potentials of the two conductors. We now connect the two conductors with a silk wire soaked in salt water. Initially, there is a potential difference between the ends of the conducting wire $\phi_2 - \phi_1$, as shown in Fig. 5.1. As a consequence, an electric field \mathbf{E} exists inside the wire (in general, it is position dependent). This is a not static situation; the system is not in equilibrium. The electric field will push the negative charges from conductor 1 to conductor 2 until the field itself, and with it, the potential difference, is zero. During this phenomenon, which is very short in practice, an electric current flows through the wire.

Consider an arbitrary section S of the wire. The current intensity through S is defined as being the electric charge going through S in a second. To be more

Fig. 5.1 Two metallic conductors connected by a conductive wire



precise, let us choose a positive direction of the wire, for example, from left to right. Let ΔQ be the charge through S in the time Δt . In any conductor, the charge is carried by the corpuscles that we classified as free charges in electrostatics, namely electrons, ions or holes. These are the charge carriers, or simply carriers. If the carriers carry positive charges, ΔQ will be positive if they move in the direction we have chosen as positive, and negative if they move in the opposite direction. If the carriers have a negative charge, as electrons do, ΔQ will be positive if they move in the negative direction and vice versa. Obviously, if both types of carriers are present (as in a semiconductor), we must consider the algebraic sum of both contributions. We then define the current intensity (with reference to the section S under consideration), the quantity I , as

$$I = \lim_{\Delta t \rightarrow 0} \frac{\Delta Q}{\Delta t} = \frac{dQ}{dt}. \quad (5.1)$$

Note that a measurement of the current intensity cannot determine the sign of the charge of the carriers, because we do not know the direction of their motion. Note also that ΔQ is the sum of an enormous number of quantized charges of the carriers. Consequently, the limit in Eq. (5.1) must be understood in the sense that the time interval Δt should become very small compared to the macroscopic times, while remaining large enough to continue to have a large number of carriers passing through in that interval.

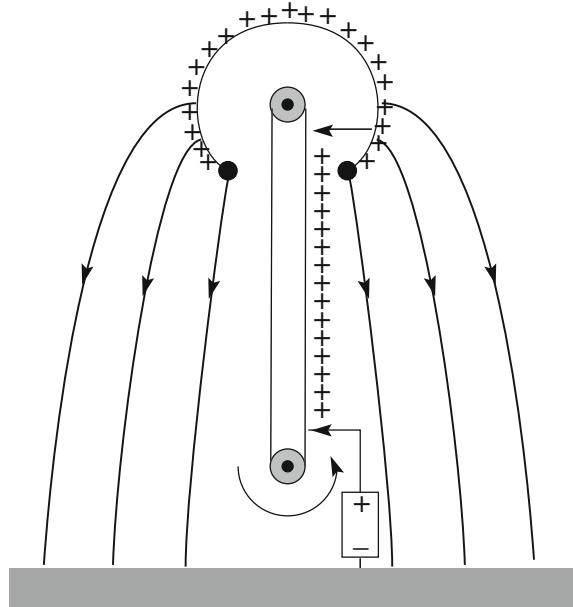
In the SI, current intensity is chosen as one of the base units. The unit is the ampere (symbol A). Its definition is based on the magnetic effects of the currents and will be given in Sect. 6.12. We recall that the unit of charge is the coulomb, which is the charge carried by a current of 1 A intensity in 1 s.

In the above example, the potential difference between the conductors $\phi_2 - \phi_1$ vanishes in a very short time, and the current with it. If we want a steady regime, namely the current intensity to be constant in time, we must keep the potential difference between the ends of the wire constant as well. A current of intensity constant in time is called a direct current. The electric field inside a conductor carrying a current cannot be zero.

Any device capable of continuously restoring the potential difference $\phi_2 - \phi_1$, which the current tends to nullify, must bring the charges arriving from conductor 1 back to conductor 2. Such devices are called generators. As generators necessarily act *against* the electric field forces, they must employ non-electrostatic forces. This is the case that we already mentioned in Sect. 1.6 of the Volta pile and of the common batteries, in which chemical processes are used. A generator employing a mechanical force was developed by Robert Van de Graaf (USA, 1901–1967) in 1929 to generate high voltages for the purpose of accelerating ions in physics laboratories. The device is not suitable for use in electric circuits, but is useful for explaining the functioning principles.

The Van de Graaf generator contains a flexible dielectric belt running on two rollers moved by a motor (Fig. 5.2). A comb of pointed electrodes faces the belt near the lower roller in the figure. These electrodes have a small positive potential,

Fig. 5.2 The Van der Waals generator



generated, for example, by a battery. The high electric field present near the tips extracts charges from the electrodes, which then flow onto the nearby running belt. The belt transports the charges to the higher roller inside a hollow metal sphere. Here, a second conducting comb is located, connected to an inner point of the sphere. The comb extracts the charges from the belt and brings them onto the sphere. The potential of the sphere grows as a consequence. Note that the charge density on the internal surface remains zero. New charges are free to join.

Clearly, the charges on the sphere generate an electric field between the sphere itself and ground. The charges on the belt move against the force of this field. The motor moving the belt produces the necessary work. We see that a force of non-electrostatic nature is employed.

Let us now go back to the electric current. The definition of current intensity we gave is suitable for conducting wires, which have a well-defined section. However, the current can flow in an extended conductor of any shape, as, for example, water does in an irregular river or lake. Consequently, we define the *current density*, a vector quantity that we call \mathbf{j} . Its magnitude is the current intensity per unit area perpendicular to the current flow. Its direction is the direction of the motion of the charges, if they are positive, and the opposite if they are negative.

Consider then a small surface ΔS normal to the motion of the carriers. If ΔI is the current intensity through ΔS , then $\Delta I = j \Delta S$. If, more generally, the surface ΔS has an arbitrary orientation, we must consider its projection perpendicular to the flow. If \mathbf{n} is the unit vector normal to ΔS , we have

$$\Delta I = \mathbf{j} \cdot \mathbf{n} \Delta S. \quad (5.2)$$

To obtain the current intensity through any extended surface S , taking into account that, in general, \mathbf{j} is a function of the point, we integrate Eq. (5.2) on the surface, giving us

$$I = \int_S \mathbf{j} \cdot \mathbf{n} dS. \quad (5.3)$$

The measurement unit of the current density is the ampere per square meter, $A m^{-2}$.

An instrument for measuring the current intensity, the galvanometer, will be studied in Sect. 6.5.

We shall now establish the relations between the motion of the carriers and current intensity and density. To fix the ideas, let us assume the carriers to have a positive charge. Consider a small surface ΔS inside the conductor and choose a positive orientation for this surface. Let \mathbf{n} be the positive normal unit vector. Let ΔI be the current intensity through ΔS in its positive direction. The surface ΔS is crossed at any moment by charge carriers. In absence of an electric field, the numbers of carriers crossing ΔS in any time interval in both directions, in their chaotic motion, are equal. The current ΔI is zero. In the presence of an electric field, more carriers cross ΔS in one of the directions than in the opposite, and ΔI is different from zero.

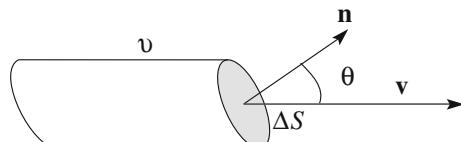
To evaluate ΔI , we should consider that the carriers have different velocities. However, as a first step, let us start by assuming all the velocities to be equal. We shall then move to considering the actual situation.

Let n_p be the number of carriers per unit volume, q the charge of each of them (in practice, the elementary charge or a small multiple) and \mathbf{v} their velocity. In general, \mathbf{v} will be at an angle, which we indicate with θ , with the unit vector \mathbf{n} normal to ΔS (see Fig. 5.3). The current intensity through ΔS is the product of q times the number of carriers crossing ΔS in a second, which we call Δn_p . In a second, a carrier moves across a distance equal to the magnitude of its velocity, v . Hence, the carriers crossing ΔS in a second are those contained in the oblique cylinder of height v in the direction of \mathbf{v} , as shown in Fig. 5.3. Hence, we have $\Delta n_p = n_p v \Delta S \cos \theta$ and $\Delta I = q n_p v \Delta S \cos \theta = q n_p \mathbf{v} \cdot \mathbf{n} \Delta S$.

Comparing this with Eq. (5.2), we see that

$$\mathbf{j} = q n_p \mathbf{v}.$$

Fig. 5.3 A section of a conductor and the velocity of a carrier



The current density is equal to the carrier velocity times their charge times their number density.

Let us now consider the real case in which every carrier has a different velocity. Let dn_p be the number of carriers per unit volume with (vector) velocity between \mathbf{v} and $\mathbf{v} + d\mathbf{v}$. These dn_p carriers have approximately the same velocity in magnitude and direction. Let dI be their contribution to the current. This is clearly $dI = qdn_p \mathbf{v} \cdot \mathbf{n} \Delta S$. We now find the total current ΔI by integrating the contributions of the carriers of different velocities. We have

$$\Delta I = \int dI = q \Delta S \mathbf{n} \cdot \int \mathbf{v} dn_p,$$

where, on the right-hand side, we have taken the constants out of the integral. Let us now recall that, by definition, the mean velocity is

$$\langle \mathbf{v} \rangle = \frac{1}{n_p} \int \mathbf{v} dn_p.$$

We conclude that

$$\Delta I = q n_p \langle \mathbf{v} \rangle \cdot \mathbf{n} \Delta S, \quad (5.4)$$

from which we have the relation between the mean carrier velocity and the current density:

$$\mathbf{j} = q n_p \langle \mathbf{v} \rangle. \quad (5.5)$$

Equation (5.5) holds for both signs of the carriers. If q is positive, the current density and mean velocity have the same sense; if q is negative, they have the opposite sense. As we already noticed, a measurement of current intensity does not distinguish between negative carriers moving in one direction and positive carriers moving in the opposite one. In Sect. 6.3, we shall see how the Hall effect in a magnetic field is capable of determining the sign of the carrier. If both positive and negative carriers are present, as is the case with semiconductors and electrolytes, the current density is

$$\mathbf{j} = |q_+| n_+ \langle \mathbf{v}_+ \rangle - |q_-| n_- \langle \mathbf{v}_- \rangle = q_+ n_+ \langle \mathbf{v}_+ \rangle + q_- n_- \langle \mathbf{v}_- \rangle, \quad (5.6)$$

with an obvious meaning for the symbols. Under the action of the field, the positive carriers move in the direction of the field, and the negative ones in the opposite direction, carrying opposite sign charge. Hence, the two contributions add on.

As in a gas, the mean kinetic energy of the thermal motion is proportional to the absolute temperature. More precisely, the mean kinetic energy per degree of freedom is equal to $1/2 k_B T$, where k_B is the Boltzmann constant and T is the absolute

temperature. In the usual situations, the carriers can be considered as point-like. The root mean square (r.m.s.) value of the velocity is

$$v_{\text{rms}} = \sqrt{\langle v^2 \rangle} = \sqrt{2\langle E_K \rangle/m}$$

Let us look at the orders of magnitude, remembering that, at room temperature, $k_B T$ is about 1/40 eV. Let us evaluate the rms velocity of monoatomic ions in an electrolyte. Considering them to be point-like, their mean kinetic energy is $\langle E_K \rangle = 3k_B T/2$. When dealing with microscopic particles, it is often useful to consider their rest energy mc^2 (where c is the speed of light) in place of the mass m and measure all energies in electronvolt. The rest energy of a nucleon being, in round figures, 1 GeV, the rest energy of a nucleus of atomic number A is about A GeV. In a sufficient approximation, the rest energy of an atom, or of an ion, is equal to the rest energy of its nucleus. For example, the rest energy of a sodium ion is about 23 GeV. Hence, we have

$$v_{\text{rms}} = c \sqrt{\frac{3k_B T}{mc^2}} = 3 \times 10^8 \sqrt{\frac{3/40}{23 \times 10^9}} \cong 540 \text{ ms}^{-1}.$$

We cannot use the same argument for electrons in a metal, because their velocity distribution obeys quantum physics and is different from the Boltzmann distribution. In any case, electrons' velocities are two orders of magnitude larger than that which we have just found for ions.

In the presence of an electric field, the mean velocity of the carriers is different from zero. It is called the *drift velocity*. Let us evaluate its order of magnitude. Consider, for example, a current of intensity $I = 1$ A in a wire having the section $S = 1 \text{ mm}^2$. Hence, the current density is $j = 10^6 \text{ A/m}^2$. We still need the number density of the conduction electrons. In general, the conduction electrons are one or two per metal ion. Let x be this number. If δ is the density of the metal, A its atomic number, and N_A the Avogadro number, we have $n_p = xN_A\delta 10^3/A$ (the 10^3 factor is because A is given in grams rather than in kilograms). Take, for example, copper. The values are $x = 1$, $\delta \cong 9 \times 10^3 \text{ kg/m}^3$, $A = 63.5$, and hence, $n_p = 0.85 \times 10^{29}/\text{m}^3$. We have then

$$\langle v \rangle = \frac{j}{qn_p} = \frac{10^6}{1.6 \times 10^{-19} \times 0.85 \times 10^{29}} \cong 7.4 \times 10^{-5} \text{ ms}^{-1}.$$

We see that the drift velocity of the conduction electrons in a metal, namely the ordered component of their motion, is very small, much smaller than their thermal velocity. The situation in semiconductors is somewhat different. The number density of the carriers (which can be positive or negative) depends strongly on temperature and on the composition of the material, on its "doping" as we say. Usually, however, n_p is much smaller than in metals. If, for example, $n_p = 10^{19} \text{ m}^{-3}$, which is not an unusual value, the drift velocity in the above considered

conditions is on the order of 10^5 m/s. In practice, however, while a metal like copper can easily “hold” a current density of 10^6 A/m² without getting too hot, this is not possible in a semiconductor. Semiconductors have a resistivity (to be defined in Sect. 5.3) much larger than that of metals and would burn out at these current densities. Reasonable values are on the order of 10 A/m² (or dozens of $\mu\text{A}/\text{mm}^2$). The corresponding carrier drift velocities are on the order of a few m/s.

QUESTION Q 5.1. Consider a current of 10 A flowing in a copper wire of 3 mm diameter. In copper there is one conduction electron for each atom. What is the average speed of the conduction electrons? □

Let us make clear, however, that the ordered motion velocities we just considered are NOT the velocities with which the changes of the current propagate in a conductor. Let us go back to the two conductors in Fig. 5.1 at the potentials ϕ_1 and ϕ_2 and consider what happens immediately after we have connected them with a wire. In the neighbor of the point at which we establish the contact, conduction electrons acquire an ordered component of motion in the direction opposite to the field. The electric field they produce immediately acts on the electrons that are located a bit farther on, which acquire a drift velocity as well. Their field, in turn, acts on still farther electrons, and so on. The development of an ordered motion propagates with a speed that is much faster than the speed of the carriers. The propagation speed is the speed of the electromagnetic waves in the medium. The following analogy is useful for understanding the mechanism. Consider a series of equal pendulums, each consisting of a metal sphere hanging on a wire. Each sphere is in contact with the one adjacent to it. If we now take the first in the row somewhat out of equilibrium and let it go, we see that, almost immediately after its collision with the second sphere, the last sphere in the row jumps up. The disturbance is propagated along the system with a much larger velocity than that of the sphere. It is the sound propagation speed in the metal of the spheres. Similarly, imagine connecting a light bulb to a battery at 1 km distance with a wire and a switch. When we close the switch, the bulb lights up immediately, after an imperceptible delay. If the bulb were to light up when the electrons coming from the generator reached it through the wire with the above-evaluated velocity, it would take about 10 years.

5.2 Electric Charge Conservation

Electric charge conservation is one of the fundamental laws of physics. It has been tested under a very large number of different conditions and has always been found to hold. While it is possible to create or destroy electric charges, their total variation must be zero. We can create a positive charge if we create an equal and opposite negative charge. For example, a photon, which is a neutral particle, can “materialize”, as we say, into an electron positron pair, or a neutron can decay into a final state consisting of a proton, an electron and a neutrino, but a neutron cannot decay into a proton and two neutrinos. Similarly, an electron and a positron can “annihilate”, producing a photon, but two electrons cannot do the same.

We can express charge conservation in formulas with the following argument. Let V be a volume enclosed by the surface S , containing the total electric charge Q . Charge conservation implies that the total charge in V can vary only by going through the surface S . Namely, the total current flowing out (in an algebraic sense) from S must be equal to the rate of decrease of the charge inside $-dQ/dt$, namely calling \mathbf{n} the normal outgoing unit vector on S :

$$-\frac{dQ}{dt} = \int_S \mathbf{j} \cdot \mathbf{n} d\mathbf{S}. \quad (5.7)$$

We can express charge conservation in a differential form, transforming the surface integral on the right-hand side into a volume integral with the Gauss divergence theorem, obtaining

$$-\frac{d}{dt} \int_V \rho dV = \int_V \frac{\partial \rho}{\partial t} dV = \int_V \nabla \cdot \mathbf{j} dV.$$

Being that this equation is true for any volume V , it must also be true between integrands, namely

$$-\frac{\partial \rho}{\partial t} = \nabla \cdot \mathbf{j}. \quad (5.8)$$

This is the *continuity equation* for an electric charge in differential form. We are stating that not only is the charge conserved, but that it is conserved *locally*. A local conservation is stronger than a global conservation. If, for example, the charge decreases by 1 C in London and contemporarily increases by 1 C in Paris, the charge is conserved globally but not locally.

A consequence of electric charge conservation is that the current intensity at any section of a circuit in a steady regime is the same. This property also holds for non-stationary regimes if the intensity variations with time are slow enough, as is often the case in practice. We shall now talk about a quasi-stationary regime under these conditions.

Let us consider a segment of a circuit, between the section S_1 upstream and S_2 downstream. Let I_1 and I_2 be the current intensities at the two sections. The currents are constant in time by assumption (steady regime). Now consider that I_1 is the charge entering the segment in a second through S_1 and I_2 is the charge flowing out of the segment in a second from S_2 . The regime being steady, the charge in the segment cannot vary. Being that the charge is conserved, it can neither be destroyed nor created. The necessary conclusion is that $I_1 = I_2$.

Similarly, the water flow through different sections of a canal is equal in a steady regime, as a consequence of the (local) conservation of mass. Contrastingly, if a dam is opened upstream and a flood wave is running through the canal, the flow in a section upstream is larger than that downstream of the wave. We can produce a similar situation in a circuit by connecting a conductive wire to the plates of a

charged capacitor. A flood wave equivalent does exist, but its velocity is enormous, on the order of the speed of light, which is 3 ns/m. The time taken to cross a circuit is consequently extremely small; it is so small that the transient can usually be considered instantaneous.

5.3 Ohm's Law

Let us connect the ends of a conducting wire to the poles of a battery providing the potential difference $V = \phi_2 - \phi_1$. Let I be the current intensity flowing through the wire. What is the relation between the applied voltage V and I ? The answer was given in 1826 by Georg Ohm (Germany, 1789–1854), who used a Volta pile in a series of experiments and established that I is proportional to V , namely that

$$V = RI. \quad (5.9)$$

This is Ohm's law, which is valid for a large class of conductors (but not all of them), which, for this reason, are called ohmic.

The proportionality constant R is the electric resistance. Its measurement unit is the ohm (Ω), which is defined as follows. A conductor has one ohm resistance if it carries a current of one ampere when it is under the potential difference of one volt. The resistance depends on the geometry of the conductor, the material of which it is made and its temperature. For wires, the resistance is directly proportional to the length l and inversely to the section S . The proportionality constant ρ depends on the material and temperature and is called *resistivity*. We have

$$R = \rho l / S. \quad (5.10)$$

The resistivity of a material is defined as the resistance of a piece of that material 1 m long with a 1 m^2 section. Its unit is the ohm times meter ($\Omega \text{ m}$). In general, resistivity depends on temperature. The dependence is not strong for metals, namely of a few per mille per degree of temperature variation. We shall come back to that at the end of the section.

Illustratively, we give examples of resistivity values at 20°C in Table 5.1.

As one can see, values of resistivity vary over the huge range of 25 orders of magnitude. There is no sharp separation between materials that can be considered good or very good conductors (like silver and copper) and good or very good insulators. Pure semiconductors are in between and get their name from that fact. Their resistivity can be increased by orders of magnitude by doping them with the proper elements. The resistivity of solutions strongly depends on their “strength”, namely the amount of solute per unit volume of solution.

As we stated, Eqs. (5.9) and (5.10) hold for wires that are ohmic conductors. If the conductor is ohmic but is not a wire, the Ohm law has a more general expression. To find it, let us consider an extended conductor, such as the one shown

Table 5.1 Resistivity in $\Omega \text{ m}$ of pure substances and a solution

Substance	Resistivity ($\Omega \text{ m}$)
Silver	1.6×10^{-8}
Copper	1.7×10^{-8}
Iron	10×10^{-8}
Graphite	$\sim 10^{-5}$
KCl in H_2O (normal solution ^a)	10^{-1}
Germanium	4.6×10^{-1}
Silicon	2.3×10^3
Boron	1.8×10^4
Iodine	1.3×10^7
Sulfur (yellow)	2×10^{15}
Quartz (fused)	$10^{16}\text{--}10^{17}$

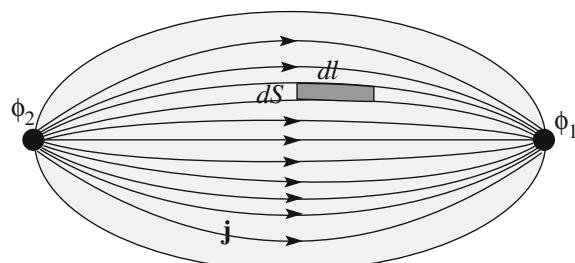
^aNormal is a solution of 1 mole of solute in 1 kg of solution (71 g/kg for KCl)

in Fig. 5.4, which is a metal sheet with two points maintained to different potentials ϕ_1 and ϕ_2 . The lines in the figure are both electric field and current density lines. Indeed, the two vectors, \mathbf{E} and \mathbf{j} , are parallel everywhere. The field lines coincide with the current lines.

Consider an infinitesimal surface dS normal to \mathbf{j} . The set of current lines crossing dS make an infinitesimal flow tube. The lateral surface of the tube is geometrical rather than physical, but still no charge exits through them because they are parallel to the velocity by construction. The flow tube behaves as a wire conductor, for which Eqs. (5.9) and (5.10) hold, with the warning that the section may vary along the tube. In order to deal with a definite section, let us consider a segment of the tube of infinitesimal length dl . Let dR be the resistance of the segment, $dl = jdS$ the current intensity through dS , and $d\phi$ the potential difference between its ends. Ohm's law gives us $d\phi = dR \cdot dl$. On the other hand, the magnitude of the electric field in the segment is $E = d\phi/dl$. In addition, Eq. (5.10) gives us $dR = \rho \cdot dl/dS$ and we can write $Edl = \rho \frac{dl}{dS} j dS$, namely $E = \rho j$. The latter relation holds for vectors too, given that they are parallel. We finally have

$$\mathbf{E} = \rho \mathbf{j}. \quad (5.11)$$

Fig. 5.4 Field lines in an extended conductor



This is Ohm's law for a conductor of arbitrary shape. It includes both Eqs. (5.9) and (5.10) and, as opposed to them, is local. Equation (5.11) is often written in an equivalent form introducing the electric conductivity, which is simply the inverse of the resistivity, namely

$$g = 1/\rho. \quad (5.12)$$

Equation (5.11) becomes

$$\mathbf{j} = g\mathbf{E}. \quad (5.13)$$

The units for conductivity are the $\Omega^{-1} \text{ m}^{-1}$. The unit for an inverse resistance, namely $1 \Omega^{-1}$, is called 1 siemens (S), and the conductivity is given in siemens per meter (S/m). To have an idea of the orders of magnitude, let us consider a copper wire ($g = 5.9 \times 10^7 \Omega^{-1} \text{ m}^{-1}$) having the section $S = 1 \text{ mm}^2$ carrying a current of $I = 1 \text{ mA}$. The current density is $j = 10^3 \text{ A/m}^2$ and the field is $E = j/g = 1.7 \times 10^{-5} \text{ V/m}$. As we can see, this is quite a small value. As a matter of fact, this field is, for comparison, on the same order as the field produced by one electron at $r = 1 \text{ cm}$ distance. Let us check:

$$q_e/(4\pi\epsilon_0 r^2) = 1.6 \times 10^{-19}/(4\pi \times 8.8 \times 10^{-12} \times 10^{-4}) = 1.45 \times 10^{-5} \text{ V/m}.$$

Ohm's law is extremely important in practice, but is not fundamental in its character. It is a phenomenological description at the macroscopic level of microscopic phenomena. At first sight, it looks strange. Indeed, the current density is proportional to the velocity of the carriers and the electric field is proportional to the force acting on them. Namely, Eq. (5.13) states that the velocity, not the acceleration, is proportional to the force. The reason for the apparent contradiction is the disordered motion of the carriers. In their motion, the carriers are subject to extremely frequent collisions, both with one another and with impurities present in the conductor. Just after a collision, the motion of a carrier is, indeed, accelerated, but only until the next collision. If τ is the mean time between collisions and \mathbf{F} the applied force, the mean velocity gained between two collisions is

$$\langle \mathbf{v} \rangle = \tau \mathbf{F}/m. \quad (5.14)$$

This is also the mean velocity in the direction of the field if, as we assume for the moment, no statistical correlation exists between the directions before and after a collision. We say that the collision randomizes the motion completely. In this case, the mean value of the velocity after a collision is zero. We conclude that the mean velocity, also called the drift velocity, is proportional to the applied force. In practice, a single collision randomizes the motion only partially. Namely, there is some correlation between incident and outgoing directions. However, any correlation disappears after a certain small number, say x , of collisions. This does not change the above conclusion if τ in Eq. (5.14) is the randomization time, which is x times the average time between collisions.

For the ohmic conductors, the carrier number density n_p is independent of the applied field and we can write $\mathbf{j} = n_p q \langle \mathbf{v} \rangle = n_p q \tau \mathbf{F}/m$. Being that $\mathbf{F} = q\mathbf{E}$, we have $\mathbf{j} = n_p q^2 \tau \mathbf{E}/m$. Looking at Eq. (5.13), we have for the conductivity

$$\sigma = n_p q^2 \tau / m. \quad (5.15)$$

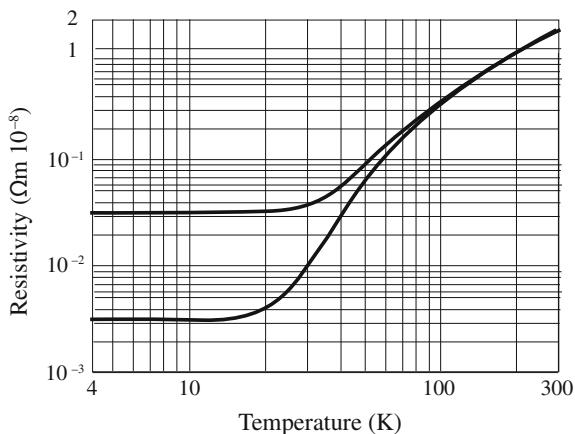
Let us discuss this equation. Firstly, it states that the conductivity is proportional to the square of the charge of the carrier. Our previous observation that current is independent of the sign of the carrier is confirmed. Secondly, we see that conductivity is larger if the randomization time is longer. Conductivity is also larger if the mass of the carrier is smaller (as for electrons compared with ions).

We can also understand why Ohm's law is not universally true. Consider, for example, an electric discharge in a gas. The carriers are some electrons and ionized molecules of the gas. If the electric field is large enough, a carrier may acquire a velocity after a collision, and a kinetic energy, large enough to break up the target molecule at its next collision. A new ionized molecule and a free electron are produced, namely two additional carriers. The process repeats itself and a chain reaction develops, producing a large number of carriers and, eventually, an electric discharge (lightning, for example). The phenomenon develops particularly in gases, in which the time between collisions is longer. In these cases, the number density n_p of carriers depends on the field and, consequently, Ohm's law does not hold.

Let us go back to the conduction in a metal. What are the targets of the electrons' collisions? One immediately thinks of the ions of the crystal. These are, however, very densely packed. The mean free path, say l , between collisions should be on the order of the size of the cell of the crystal lattice, a , which is on the order of 0.1 nm. A simple calculation leads to conductivity values two orders of magnitude smaller than those observed. Our model based on classical physics is wrong.

Let us consider the additional experimental information on the temperature dependence of the metal resistivity. This is shown in Fig. 5.5 for pure copper. The resistivity decreases with decreasing temperature, tending for $T \rightarrow 0$ to definite values that are different from zero, called the residual resistance. The curves in Fig. 5.5 are for two different metallurgic treatments. The resistivity of pure copper at the ambient temperature is $\rho_{\text{Cu}} = 1.71 \times 10^{-8} \Omega \text{ m}$. The residual resistance, measured at $T = 4 \text{ K}$, is 50–500 times smaller, depending on the treatment of the conductor. This behavior cannot be explained assuming the ions of the crystal to be the target of electron collisions. Indeed, the size of the crystal cell varies just a little and decreases, if at all, with temperature. The consequence should be a modest increase of resistance with decreasing temperature, contrary to observations. The explanation is given by quantum mechanics, which shows that electrons move in a perfect lattice, like in a vacuum. They never collide with the ions. Contrastingly, the collisions are on defects of the lattice and on impurities, namely atoms different from the crystal ones, which are always present in small amounts. This accounts for the residual resistance. When the temperature increases, the thermal motion of the ions increases. The ions oscillate about their equilibrium positions, which are the

Fig. 5.5 Resistivity versus temperature in double logarithmic scales for pure copper of different treatments



vertices of the lattice, with increasing amplitude. The probability of an electron finding an ion “out of position” increases. This explains the increase in resistivity with temperature.

5.4 The British Association Ohm

This section is a historical interlude, which can be skipped without compromising the understanding of that which follows it. It is dedicated to showing, along with an example, the importance of the standardization of measurement units. Today, standard resistance boxes are so easily commercially available as to be taken for granted. However, the labor necessary to secure the initial reliable standards was huge, requiring organization and close collaboration between engineers and physicists.

In the 1840s, a network of telegraph lines was already operational across Britain, consisting partially of underground cables, partially of cables strung between poles. No particular precision resistance measurements had been needed for such installations. The situation changed when the plans for laying the first Atlantic telegraphic cable were begun. This would be the first step of an immense commercial undertaking, ultimately leading to a world-wide telecommunications cable network. The commercial, political, cultural and military consequences were enormous. The success of the enterprise would not have been possible without the ingenious scientific input of William Thomson (UK, 1824–1907).

Thomson was appointed as the leading science consultant by the Atlantic cable company in the 1850s. Initially, however, he had been given insufficient scientific control of the process, as became evident with the failure of the first cable-laying expedition in 1858. The following year, the British government and the Atlantic Telegraph Company set up a joint committee to investigate the causes of the failure. In its final report, delivered in 1861, the committee advanced a number of

recommendations, including, on Thomson's suggestion, a call for standard resistances to be used in all future contract specifications. The next cable-laying expedition took place with the *Great Eastern* in 1866. Thomson, having gained the full confidence of the Company, had the authority to oversee all the phases of the process on board, leading to a spectacular success. But let us look at what happened in between.

Much of the merit must be credited to the development of the "British Association ohm", lead by W. Thomson, and to its official recognition in the UK as the standard of electric resistance. Accurate and reliable resistance measurements in long distance cable technology telecommunication are crucial for several reasons. Firstly, the telegraphic signal in the Atlantic cable must propagate over distances of thousands of kilometers without deterioration. In the first decennia of the 19th century, it had become clear that high transmission rates could be achieved only by keeping the copper resistance along the cable as low as possible. Even more dangerous were any changes in resistance between sections of the cable. As a matter of fact, Thomson, testing supposedly equal "pure" copper wires, had found their resistance sometimes to differ by a factor of nearly two. Further investigations showed that this was due to the presence of small quantities of impurities, depending on the metallurgic process used to produce the wire. Secondly, the contracts with the companies for industrial cable production must rely on officially recognized and legally binding standards, both for measurement and quality control. Such standards are crucial in settling possible disputes. Thirdly, resistance measurements of the cable on board of the cable-laying ship are mandatory. The cable is quite delicate and the maneuvers necessary to prepare it for deploying can produce faults in the wire or leaks in the insulating cladding, which must be detected before deploying. Accurate and standardized procedures for resistance measurement allow for detecting an anomaly and precisely localizing the fault.

In 1861, the British Association for the Advancement of Science appointed a committee for the establishment of electrical standards, including engineers and physicists, lead by W. Thomson. The Committee gave priority to resistance measurements. The task included both the design and construction of the physical prototype and the definition of procedures to measure its resistance. The standard prototype resistor must be stable over time within a specified maximum variability and be accurately reproducible in secondary standards. The measurement must be feasible within a specified accuracy and, very importantly, must be *absolute*. Namely, the measurements should not be done by comparison with other electrical standards, like those for voltage or current intensity, but by experiments leading to a comparison with the base units of mass, length and time alone. The absolute definition of the unit is necessary to make it reproducible, a necessary requirement for its establishment by law. To appreciate the importance of the absolute, rather than relative, method, suppose that, after some time, doubts might arise about the stability of the certified standard resistor. The absolute measurement would then be repeated and the result referred to the mass, length and time standards. Any possible intervening changes would be detected and measured. Note also that the resistance of the standard does not need to be exactly 1Ω , once we know exactly how to measure it absolutely.

The measurements on the prototype standard may be quite complex, requiring a dedicated laboratory and a high level of skill. The procedure for producing, and calibrating, sufficiently accurate replicas and performing measurements even in hostile environments, like on a transatlantic ship, had to be defined. After 2 years of difficult analysis and measurements, the committee finally issued its official ohm standard in February 1865.

The method for the construction of the absolute ohm is the spinning coil method. It was proposed by W. Thomson (who was knighted in 1866 and made Baron of Kelvin in 1892 for his merits) and adopted by the committee. It is a masterpiece of experimental art, which, unfortunately, we will not describe here in detail, because it is based on the electromagnetic induction phenomena that we shall study in Chap. 7. We simply mention that the current induced by the spinning of the coil in the earth magnetic field generates, in turn, a magnetic field. This field was determined by measuring the mechanical torque with a torsion balance, whose arm was a suspended permanent magnet. Only mechanical measurements were needed. The ingenious design was such that the final effect was independent of the magnitude of the local earth magnetic field. Well-known physics laws and accurate calculations allowed for the expression of the result of these measurements as a relation between the value of the resistance of the standard and the units of mass, length and time.

Once the “British association ohm” was established as the primary standard, a number of replicas for research and industrial use were produced. These secondary standards are calibrated through *relative* measurements against the primary standard.

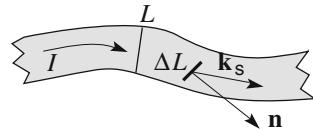
The afore-mentioned successful completion of the *Great Eastern* expedition in 1866 was soon followed by others to deploy cables to India, China, Japan and Australia, with great geopolitical consequences. The work on the British Association ohm had considerable impact on basic science as well. In 1862, James Clerk Maxwell (UK, 1831–1879) was appointed to serve on the Committee on Electrical Standards. His experimental and conceptual contributions to the program were fundamental for the development of his experiment for the absolute measurement of the ratio of the electric and magnetic constants, the final step in his development of his electromagnetic theory of light. We shall describe the experiment in Chap. 10.

5.5 Surface Currents

Consider a conductive layer of negligible thickness carrying a steady current. The current intensity I is also, in this case, the charge passing through a section of the conductor in a second. However, the section is now a line, as L in Fig. 5.6. It is sometimes called a surface current to indicate its geometry, but it is just a current measured in amperes.

The definition of the (surface) current density \mathbf{k}_s is a bit different from the case of volume currents. Let ΔL be a small, arbitrarily oriented section of the surface and \mathbf{n} its unit normal vector. The current ΔI through ΔL is

Fig. 5.6 A conductive thin layer and a surface current



$$\Delta I = \mathbf{k}_s \cdot \mathbf{n} \Delta L. \quad (5.16)$$

The vector \mathbf{k}_s is obviously tangential to the current surface. Its physical dimensions are A/m (rather than A/m^2 , as in the case of the volume current density) because the section of the conductors has the dimensions of a length (rather than an area).

Surface currents may be present, not only on thin metal sheets, but also on the surface of solid conductors. We shall encounter these in the following sections. An example is shown in Fig. 5.7, where the current flows along the surface of a cylindrical conductor.

Ohm's law also holds for the surface Ohmic conductors. Consider the two-dimensional equivalent of a constant section wire, namely a constant width conductive ribbon, as shown in Fig. 5.8.

Let V be the potential between its ends and I the current intensity. It is found that I is proportional to V ,

$$V = RI, \quad (5.17)$$

which is Ohm's law. Also, in this case, R is directly proportional to the length of the conductor l and inversely to its section L

$$R = \rho_S l / L, \quad (5.18)$$

Fig. 5.7 A cylindrical current

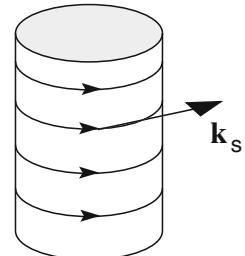
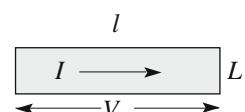


Fig. 5.8 A segment of constant section surface conductor



The only difference is that now the section has the dimension of a length. Consequently, the dimensions of the *surface resistivity* ρ_s are equal to the dimensions of R . The surface resistivity is the resistance of a square of that surface. We do not need to specify the size of the side of the square (1 m^2 , for example) because every square of the surface has the same resistance. Indeed, if, starting from a given conducting square, we double its side, its resistance does not vary, because the length of the conductor doubles (which would double the resistance) but its section doubles as well (which would halve the resistance). Consequently, the surface density measurement unit is usually called ohm per square (Ω/\square), which is just the ohm.

5.6 Energy Balance

Ohm's law states that in order to have a steady current of intensity I in a resistor of resistance R , we must apply a voltage $V = RI$ to its extremes. This implies that energy is dissipated in the resistor. This energy is delivered by the generator producing the voltage V . As we have seen, the generator does work against the electric field to carry back the charges. If dq is the charge crossing the resistor in the time interval dt , the work of the generator is Vdq . The work per unit time, namely the power delivered by the generator, is

$$w = V \frac{dq}{dt} = VI = RI^2 = \frac{V^2}{R}. \quad (5.19)$$

This is the energy dissipated in the resistor per second. The dissipated energy remains internal, i.e., the thermal energy of the conductor, increasing its temperature. In the same time, the conductor emits energy to the environment, soon reaching an equilibrium temperature at which the input electric power is equal to the power delivered to the environment. This is the case with electric heaters, electric light bulbs, etc. The effect was discovered with a series of careful experiments between 1840 and 1843 by James Prescott Joule (UK, 1818–1889) and is called Joule's law of heating or simply the Joule effect.

Equation (5.19) is expressed as a global law, regarding the entire resistor. The law can be expressed in a local form too. Let us consider an extended conductor, as we did in Sect. 5.3, carrying a steady current. Let $\mathbf{j}(\mathbf{r})$ be the current density at the point of position vector \mathbf{r} . To simplify the argument, suppose that all the carriers have the same charge q . Let n_p be the carrier number density, \mathbf{v}_i the velocity of the i -th carrier and $\mathbf{E}(\mathbf{r})$ the electric field in the conductor. The work done by the field on the i -th carrier in a second is $\mathbf{F}_i \cdot \mathbf{v}_i = q\mathbf{v}_i \cdot \mathbf{E}$. Recalling that $\mathbf{j} = n_p q\langle \mathbf{v} \rangle$, we write that the work done by the field per unit time in the unit volume is

$$\sum_{i=1}^{n_p} \mathbf{F}_i \cdot \mathbf{v}_i = q \left(\sum_{i=1}^{n_p} \mathbf{v}_i \right) \mathbf{E} = q n_p \langle \mathbf{v} \rangle \cdot \mathbf{E} = \mathbf{j}(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r}). \quad (5.20)$$

We can then state that the power dw dissipated in the elementary volume dV is

$$dw = \mathbf{j}(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r}) dV,$$

which is the local equation we have been looking for. To know the power dissipated at a point of the conductor, we just have to know the electric field and current density at that point.

5.7 Generators

We shall now consider continuous voltage generators and their characteristics as circuit elements. We start with the simplest configuration, consisting of a resistor connected to a generator, for example, a battery. Let us consider a closed line completely inside the circuit and having the direction of the current density at each point. By construction, the integral over that line of the current density is positive, namely $\oint \mathbf{j} \cdot d\mathbf{s} > 0$. Let us assume Ohm's law, namely $\mathbf{j} = g\mathbf{E}$, to hold everywhere in the circuit. Being that g is positive, we also have $\oint \mathbf{E} \cdot d\mathbf{s} > 0$. But this cannot be under static conditions in which the field \mathbf{E} is conservative. We must conclude that the equation $\mathbf{j} = g\mathbf{E}$ does not hold, at least in part of the circuit. At least in a segment of the circuit, the current must be opposite to the electric field. As we know, this happens inside the generator. For example, in the Van de Graaf generator, it is the belt that exerts a force on the charges, pushing them against the field force. In the Volta pile, the forces opposite to the electrostatic field have a chemical origin.

We now define a quantity that should not be confused with the potential difference, even if it has the same physical dimensions and is measured in volt. This quantity is called the electromotive force (emf, for short), for historical reasons, but is clearly not a force. An emf refers to a segment of a circuit, say between points A and B . The electromotive force is the line integral on the circuit element from A to B of the acting force per unit charge.

$$\text{emf} = \int_A^B \frac{\mathbf{F}}{q} \cdot d\mathbf{s}. \quad (5.21)$$

The electromotive force is the main characteristic of a generator, as a component of an electric circuit. It is the work done by the generator on the unit charge to move it inside from the negative to the positive pole.

Let us see how to measure the emf of a generator. We connect a resistor, of resistance R , to its poles, as shown in Fig. 5.9. The generator is represented by the gray area. The two lines, one longer and thinner, one shorter and thicker, are the standard symbol for an emf generator. The generator is characterized not only by its emf \mathcal{E}_0 , but also by its resistance. It is called *internal resistance* and is denoted by R_i . We assume Ohm's law to be valid for the generator as well.

Consider the unit charge moving along the entire circuit. As energy is conserved, the work done by the generator on the charge must be equal to the energy dissipated on the resistances. There are two of them, the resistor and the internal resistance of the generator. We then have $\mathcal{E}_0 = IR + IR_i$ and hence,

$$I = \frac{\mathcal{E}_0}{R + R_i}. \quad (5.22)$$

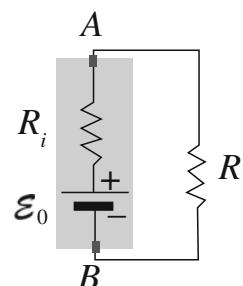
The potential difference across the external resistor R is

$$V = RI = \frac{\mathcal{E}_0}{1 + R_i/R}, \quad (5.23)$$

which is smaller than the emf of the generator. V is the voltage we measure with a voltmeter between the poles under these conditions. We see that $V \rightarrow \mathcal{E}_0$ for $R \rightarrow \infty$. Now, an infinite resistance exists when the circuit is open. This means that the emf of the generator is the potential difference between its poles measured at open circuit. When we close the circuit, the potential difference between the poles decreases, the more so the smaller R is relative to R_i . This implies that part of the energy delivered by the generator is lost inside the generator and is not externally available for R . The latter is called a load and might be, for example, an electric heater. An ideal emf generator should have zero internal resistance, but does not exist.

In the above case, we discussed the emf being localized between the poles of the generator. We shall see that, under dynamic conditions, the emf can be distributed along the entire circuit.

Fig. 5.9 A emf generator closed on a resistor



5.8 Slow Capacitor Discharge

In this section, we consider examples of non-stationary conditions. The circuit in Fig. 5.10 consists of a capacitor, a resistor and a switch. Initially, the switch is open and the capacitor is charged with the charge Q_0 . At time $t = 0$, we close the switch. The capacitor's charge $Q(t)$ and voltage $V(t)$ will vary in time. We want to find these functions, which are linked by the relation $V(t) = Q(t)/C$.

In circuits of usual sizes, the current intensity can be safely considered to be the same in every section, because its variations propagate at speeds close to the speed of light. For example, in round numbers, for a circuit of 10 cm diameter, the propagation time of the current variations is about one nanosecond. For Ohm's law, the current intensity at time t is then $I(t) = -dQ(t)/dt = V(t)/R$. Notice the minus sign, which is there because a positive current corresponds to a decreasing charge of the capacitor. Putting things together, we find that the function $Q(t)$ obeys the differential equation

$$\frac{dQ}{dt} + \frac{1}{RC}Q = 0, \quad (5.24)$$

which is solved by separating variables, obtaining

$$Q(t) = Q_0 e^{-t/RC}, \quad (5.25)$$

where we have imposed the initial condition $Q(0) = Q_0$. The charge of the capacitor (and consequently, its potential difference) decreases exponentially with time. It is easy to see that the quantity $\tau = RC$ has the dimensions of time. It is called the *time constant* of the circuit. The time constant is the time in which the charge of the capacitor diminishes by a factor $1/e$. The potential difference is

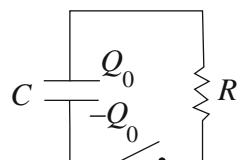
$$V(t) = V_0 e^{-t/RC}, \quad (5.26)$$

where, obviously, $V_0 = V(0) = Q_0/C$.

Let us now calculate the energy dissipated on the resistor. In the elementary time interval dt , it is

$$dW = \frac{V^2}{R} dt = \frac{V_0^2}{R} e^{-2t/RC} dt.$$

Fig. 5.10 Discharging a capacitor on a resistor



The total dissipated energy is then

$$W = \frac{V_0^2}{R} \int_0^{\infty} e^{-2t/RC} dt = \frac{CV_0^2}{2},$$

which is, as it should be, the initial energy of the capacitor.

Let us now look at the charging process of a capacitor. Figure 5.11 represents a circuit containing a generator of emf \mathcal{E}_0 , a resistor, a capacitor and a switch. The charge and the potential difference of the capacitor are initially zero. At time $t = 0$, we close the switch. Let $V(t)$ be the voltage of the capacitor in the generic instant, $Q(t)$ its charge and $I(t)$ the current intensity in the circuit.

The potential difference on the resistor is $\mathcal{E}_0 - V(t)$. Hence, for Ohm's law, we have $\mathcal{E}_0 - V(t) = RI = R \frac{dQ}{dt}$, which can be written as $R \frac{dQ}{dt} + \frac{Q}{C} - \mathcal{E}_0 = 0$.

Clearly, at the end of the transient, the current intensity will be zero and the voltage across the capacitor will be \mathcal{E}_0 . Calling Q_0 the final charge of the capacitor, $Q_0 = \mathcal{E}_0 C$, we can re-write the equation as $\frac{dQ}{dt} + \frac{1}{RC}(Q - Q_0) = 0$, and finally, being that Q_0 is a constant, as

$$\frac{d(Q - Q_0)}{dt} + \frac{1}{RC}(Q - Q_0) = 0, \quad (5.27)$$

which is solved by separating variables, obtaining

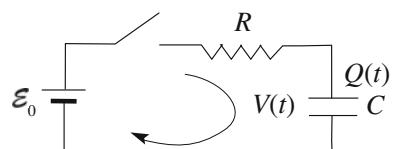
$$Q = Q_0 \left(1 - e^{-t/RC}\right), \quad (5.28)$$

where we have imposed the initial condition $Q(0) = 0$. We see that the charging is also an exponential process, with the same time constant RC as the discharge.

QUESTION Q 5.2. A $10 \mu\text{F}$ capacitor is charged at 220 V and then connected to a $10 \text{ k}\Omega$ resistor. How long would it take to the capacitor to "fully discharge" itself, if we define that as five time constants? How much will be voltage, charge and energy at that moment? \square

QUESTION Q 5.3. A $10 \mu\text{F}$ capacitor discharges through a $10 \text{ k}\Omega$ resistor. How much are the voltage across the capacitor and the current intensity after 50 ms ? \square

Fig. 5.11 Charging a capacitor



5.9 Circuits in a Steady Regime

We shall now consider electric circuits made of a number of resistors and generators in a steady regime. Two are the basic connections of circuit elements, resistors in this case, in series and in parallel.

If two or more resistors are connected in series, they carry the same current, as shown in Fig. 5.12.

Let us consider, for simplicity, two resistors (although the argument is valid for any number). Calling V_A , V_B , and V_C the potentials in the corresponding points in the figure, I the current intensity and R_1 and R_2 the two resistances, Ohm's law gives $V_B - V_A = R_1 I$ and $V_C - V_B = R_2 I$. Adding the two equations, we have $V_C - V_A = (R_1 + R_2)I$. Namely, two (or any number of) resistors in series behave the same as a single resistor with resistance equal to the sum of the resistors in series.

$$R_{\text{series}} = \sum R_i \quad (5.29)$$

Two resistors are connected in parallel when the potential difference at their ends is the same. Let us consider the case of two resistors, again with a generally valid argument, as shown in Fig. 5.13.

The current intensities in the two resistors are $I_1 = (V_B - V_A)/R_1$ and $I_2 = (V_B - V_A)/R_2$. Hence, the total intensity is $I = I_1 + I_2 = (V_B - V_A)(R_1^{-1} + R_2^{-1})$. Two resistors in parallel are equivalent to a single resistor with resistance

$$R_{\text{parallel}} = (R_1^{-1} + R_2^{-1})^{-1} = \frac{R_1 R_2}{R_1 + R_2} \quad (5.30)$$

or, for any number,

$$R_{\text{parallel}} = \left(\sum R_i^{-1} \right)^{-1}. \quad (5.31)$$

The rules set out in Eqs. (5.29) and (5.30) can be used to reduce more complicated networks, for example, the one shown in Fig. 5.14.

Fig. 5.12 Resistors in series

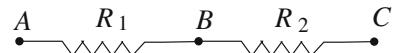


Fig. 5.13 Two resistors connected in parallel

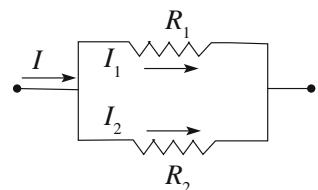
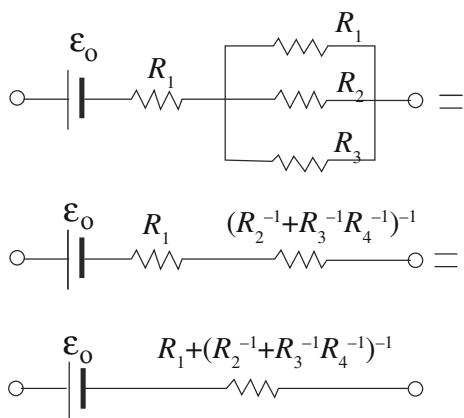


Fig. 5.14 Reducing a network of resistors to a single equivalent one



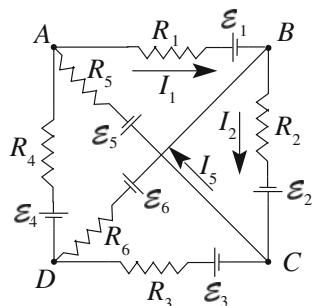
However, networks exist, such as the one in Fig. 5.15, which cannot be solved with these simple rules. With reference to Fig. 5.15, the problem is finding the current intensities I_1, I_2, \dots, I_n in each of the sides of the network, knowing the values of the emfs and of the resistances. These problems are solved using two rules, introduced in 1845 by Gustav Kirchhoff (Germany, 1824–1887) and called the Kirchhoff rules. The analysis of the circuit starts (arbitrarily) with the choice of a positive sign for each branch of the circuit. We then identify a number of *nodes* and a number of *loops* or *meshes*. A node is a point at which two or more prongs join, for example, points *A* and *B* in the figure. A mesh, or loop, is a set of prongs making a closed circuit, such as *ABCD*, *ABC*, or *BDA* in the figure.

The nodes rule, or the first Kirchhoff law, says that the algebraic sum of the current intensities converging in a node, say I_1, I_2 , taken with a positive sign if they enter the node, negative if they exit, must be zero. This is just charge conservation. Namely,

$$\sum_k I_k = 0. \quad (5.32)$$

There is one such equation for every node.

Fig. 5.15 Example of non-reducible resistor mesh



The second Kirkhoff law applies to the loops. Consider, for example, the loop $ABCA$ and choose one of the two orientations as positive, for example, the clockwise one. We give the positive sign to the currents running in the positive orientation of the loop and the negative to the others. We give the positive sign to the emfs generating currents in the positive orientation and vice versa to the other. Then, we apply Ohm's law. In this way, we obtain, for example, for the loop $ABCA$, which has three prongs, the three equations

$$R_1 I_1 = V_A - V_B - \mathcal{E}_1$$

$$R_2 I_2 = V_B - V_C + \mathcal{E}_2$$

$$R_3 I_3 = V_C - V_A - \mathcal{E}_5.$$

Adding these equations, the potential differences cancel each other out, and we get

$$R_1 I_1 + R_2 I_2 + R_3 I_3 = -\mathcal{E}_1 + \mathcal{E}_2 - \mathcal{E}_5.$$

The conclusion is general, being a consequence of energy conservation. The rule is that the algebraic sum of all the emfs of the generators in the loop (namely the energy they deliver on the unit charge going around the loop) is equal to the sum of the voltage drops RI in all the branches of the loop, namely

$$\sum_{i=1}^n R_i I_i = \sum_{i=1}^n \mathcal{E}_{1i}. \quad (5.33)$$

where n is the number of prongs of the loop.

The number of equations obtained applying the two Kirkhoff rules to all the nodes and all the loops of the network is larger than the number of unknown, so that not all are needed. Some experience is enough to learn the most convenient choices.

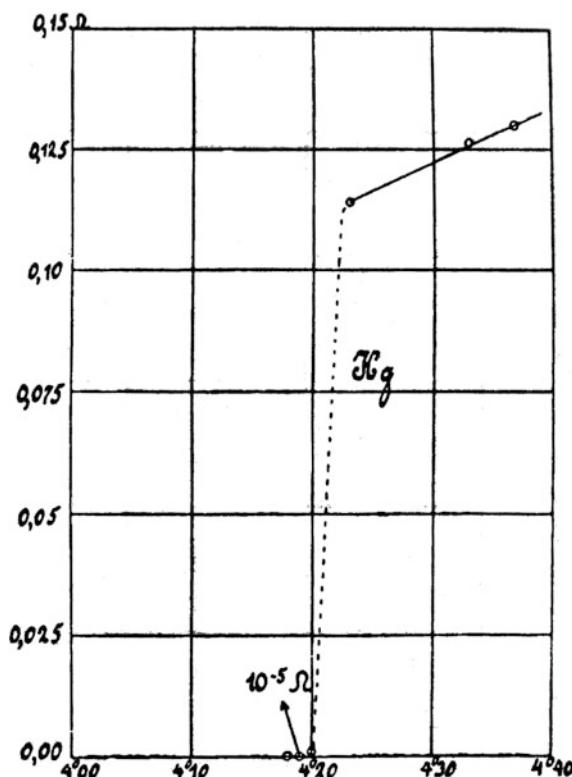
5.10 Superconductivity

Superconductivity is the phenomenon of zero electrical resistance characteristic of a number of substances below a characteristic transition temperature. In 1908, Heike Kamerlingh Onnes (The Netherlands, 1853–1926), working in the cryogenic laboratory he had established in Leiden, succeeded in liquefying helium at 4 K, a temperature that nobody had reached before. Kamerlingh Onnes dedicated the following few years to developing, with the help of his technician Gerrit Flim, an advanced cryostat designed for conducting experiments using liquid helium as a refrigerator (after having transferred the liquid from the liquefaction dewar). The temperature was measured with a gas thermometer and could be lowered below 4 K by reducing the vapor pressure of the helium.

The cryostat was ready in April 1911. Already in the first experiment, on April 8, Onnes had gold and mercury resistors (liquid mercury solidifies at 234 K) in the cryostat. After having accurately measured the gold and mercury resistance at 4.3 K, the team started to decrease the helium vapor pressure in the cryostat. Gold and mercury resistance were measured again when the temperature was 3 K. The latter was “practically zero”, as Onnes wrote in the logbook. The mercury resistance remained immeasurably small down to the minimum temperature of 2.2 K reached that day. Figure 5.16 is the historic plot, obtained a few months later in October 1911, of the resistance of the mercury resistor (in ohm) as a function of temperature (in kelvin). Measurements were done by slowly increasing temperature, a procedure allowing for a better control of temperature than when decreasing it. The plot shows the superconducting transition at the transition temperature $T_c = 4.20$ K. Within 0.01 K, the resistance jumps from immeasurably small values ($<10^{-6} \Omega$, the sensitivity limit) to 0.13 Ω .

A few months later, the Leiden team discovered that lead and tin were also superconductors, with transition temperatures of 7.2 and 3.7 K, respectively. A far greater number of superconductors with cryogenic transition temperatures, namely below 10 K, are known today.

Fig. 5.16 Onnes' measurement of the resistance of a mercury (Hg) resistor (in ohm) versus absolute temperature (in kelvin), showing the transition to the superconducting phase, for which the upper limit of resistance is reported



More than half a century later, in 1986, Johannes Georg Bednorz (Germany, born 1950) and Karl Alexander Müller (Switzerland, born 1925), discovered that a lanthanum-based compound, a cuprate peroxide to be precise, had a surprisingly high transition temperature at $T_c = 35$ K. It was soon found that replacing lanthanum with yttrium raised the transition temperature to $T_c = 92$ K. This was already higher than the liquid nitrogen temperature at atmospheric pressure, which is 77 K. Liquid nitrogen being cheap and easily available, this makes “high temperature” superconductivity much easier to study. Compounds with even higher transition temperatures were found in the subsequent years.

Superconductivity is a genuine quantum phenomenon. We shall, however, give some hints to provide an intuitive idea as to how the phenomenon works, limited to low temperature superconductivity. Its theoretical interpretation was given by John Bardeen (USA, 1908–1991), Leon Cooper (USA, born 1930) and John Robert Schrieffer (USA, born 1931) in 1957.

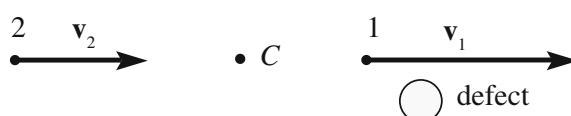
Below the transition temperature, the carriers in a superconductor are not single electrons, but correlated pairs of electrons, called Cooper pairs. The pair is linked by a tiny force, of quantum origin. The binding energy of the pair is only of a few meV. If the temperature is low enough, the available thermal energy is insufficient for breaking the pair.

This attractive force is due to the crystal lattice. The electric field of one of the two electrons slightly modifies the crystal (attracting the positive ions closer to it) and the second electron is attracted towards the resulting region in which there is a small excess of positive charge. Note that the ions need some time to go back to their unperturbed position, due to their inertia. Consequently, the second electron may be located at a sizeable distance from the first one. The mean distance between the two electrons of the pair, called the *correlation distance*, is on the order of 100 nm. This is very large compared to the cell sides, which are on the order of 100 pm.

The quantum laws governing the system are such that when the current is zero, the electrons of the pair have equal and opposite linear momenta. Consequently, the velocity of their center of mass is zero. When a field is present, the pair moves in the direction opposite to that of the field. Let us see what happens when a member of the pair approaches a defect in the crystal. As we know, under the usual conditions, resistivity is due to the scattering of electrons off the impurities. In a superconductor, when one member of the pair is close to the defect, its partner is quite far from it. This is shown schematically in Fig. 5.17.

The center of mass, C in the figure, moves with velocity $(\mathbf{v}_1 + \mathbf{v}_2)/2$. In the center of mass reference, the two electrons have equal and opposite velocities. In order to scatter, the electron close to the defect should change its velocity. But this

Fig. 5.17 A cooper pair and a crystal defect near to one partner. Point C is the center of mass of the pair



velocity is correlated with that of the partner, which is far. Namely, a scattering would imply breaking the pair. But this cannot happen, because the energy is not available. In conclusion, there is no scattering, the motion does not get disordered, and there is no resistivity.

When temperature increases, a growing fraction of the pairs breaks down, absorbing thermal energy. Then, both the Cooper pairs contributing to superconductivity and the unpaired electrons contributing to normal conduction are present. At the critical temperature, no Cooper pair is left, and the conductor becomes normal.

Superconductors have important practical applications. Even though, by 1914, Onnes had already discovered that superconductivity is destroyed by the presence of a magnetic field of intensity above a material dependent critical value, techniques were later developed, and are still being developed, to produce superconducting electromagnets generating high magnetic fields. The high intensity currents, which are necessary to produce those fields, flow through the superconductor under zero emfs and without any Joule energy loss. In practice, the superconductive circuit is in a liquid helium bath. Nowadays, the 8.3 T bending magnets of the Large Hadron Collider at CERN use cables of a niobium-titanium alloy, whose critical magnetic field is large enough. The entire magnetic structure is immersed in a 27 km liquid helium circuit. Superconductive magnets are used in high resolution NMR for medical diagnostic purposes, in mass spectrometers, etc.

Summary

In this chapter, we have learned the following principal concepts:

1. The current intensity (a scalar quantity) and the current density (a vector quantity) and their relations with the motion of carriers
2. The local conservation of electric charge
3. Ohm's law, holding for a large class of materials, and its microscopic interpretation
4. The energy balance in electric current transport
5. The electromotive force and its generators
6. The analysis in direct current of circuits composed of loops of resistors and generators.

Problems

- 5.1. Can the field lines of \mathbf{j} be closed in a steady regime? Can they radiate from a point? Can that happen in an arbitrary regime?
- 5.2. First, connect a voltmeter to a battery, then disconnect it and connect it to another battery. The readings are different. Why?

- 5.3. At the point of coordinates x, y, z of a conducting medium, the current density is $\mathbf{j} = 3x^2y \mathbf{u}_x - 3xy^2 \mathbf{u}_y + xy \mathbf{u}_z$ (where $\mathbf{u}_x, \mathbf{u}_y, \mathbf{u}_z$ are the unit vectors of the axes). Is the charge density at that point constant or variable in time?
- 5.4. We have five equal resistors. We connect two of them in series and two of them in parallel. We then separately connect to the poles of a generator: (a) the resistor alone, (b) the pair in series, (c) the pair in parallel. In which case is the power absorbed at a maximum? (The internal resistance is negligible.)
- 5.5. State whether the resistivity of an ohmic material does or does not depend on: the geometrical dimensions, the temperature, the material, and the current intensity.
- 5.6. When a bolt of lightning discharges an intense current between a cloud and the earth (or between two clouds), the current exists through the atmosphere. Does Ohm's law hold with the atmosphere resistivity?
- 5.7. A string of Christmas lights is made up of 50 bulbs in series. One of them burns out. You take it out and restore the connection of the remaining 49. In what case is the total light of the bulbs higher, if you connect the string to the same battery in both cases?
- 5.8. Figure 5.18 shows a section of a circuit (between points A and B) carrying the continuous current of intensity I in the shown direction. The upper part of the figure shows the voltage drops along the first elements of the section. The segment drawn inside the generator is schematic. Complete the diagram. Why does the first generator give a voltage drop, rather than a voltage increase? Is the same true for the other ones?

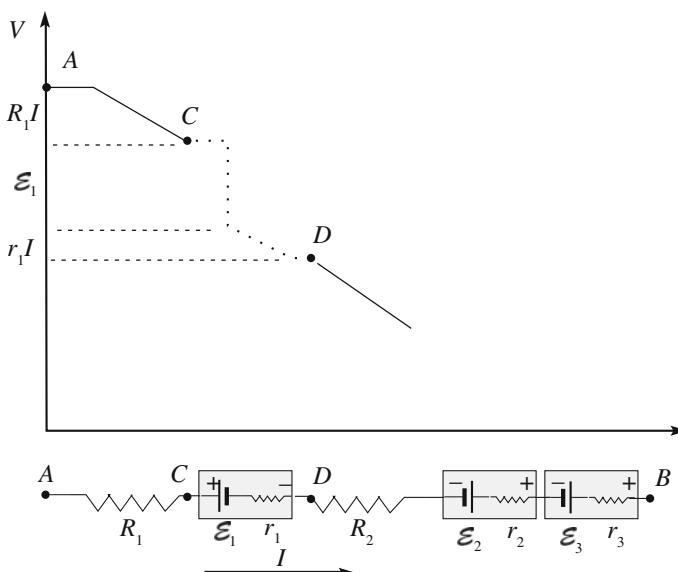


Fig. 5.18 A circuit section and the voltage across it

- 5.9. A 12 V battery is connected to a $10 \text{ m}\Omega$ resistor. What is the current intensity (before you damage the battery)? Or are the data insufficient to answer?
- 5.10. Two batteries with the same emf E and different internal resistance R_1 and R_2 are connected to the same resistance R . Can the value of R be chosen in order that the voltage difference between the poles of the first battery be zero?
- 5.11. Five identical batteries have the same emf $E = 1.5 \text{ V}$ and internal resistance $R_i = 0.2 \Omega$. We connect them first in series, second in parallel. What are the equivalent emfs and internal resistances in the two cases?
- 5.12. We have a Ni–Cr wire (resistivity equal to $10^{-6} \Omega \text{ m}$) of 1 mm^2 section. With it, we want to build a cooker dissipating 100 W. How long should the wire be?
- 5.13. A battery with emf $E = 1.5 \text{ V}$ and internal resistance $R_i = 1 \Omega$ is connected to a resistor of variable resistance R from 1 to 10Ω . Draw the diagram of the potential difference across the resistor as a function of its resistance.
- 5.14. A parallel plate capacitor has square plates with sides of $l = 250 \text{ mm}$ separated by a distance $h = 3 \text{ mm}$. The plates are vertically arranged and connected to a battery of emf $V = 200 \text{ V}$. Initially, the capacitor is surrounded by air, then we lower it into an oil bath ($\kappa = 5$) with a speed of $v = 2 \text{ mm/s}$. What is the current intensity delivered by the battery (neglecting the internal resistance)?

Chapter 6

Magnetostatics

Abstract In this chapter, we study magnetic phenomena under time-independent conditions. We define the magnetic field \mathbf{B} on the basis of the force acting on an exploring charge and see how \mathbf{B} can be measured exploiting the Hall effect. We discuss examples of motion of charged particles in magnetic fields. We then show how electric currents generate magnetic fields, find the corresponding differential equations and solve them. We give the definition of the ampere. The concept of vector potential is introduced and discussed. We then deal with the properties of the magnetic dipole. We finally discuss the transformation properties of charge and current densities between inertial frames in relative motion.

In this chapter, we study magnetic phenomena under time-independent conditions. Ancient civilizations long ago knew the property of certain ferrous minerals to attract pieces of iron. Lodestone, the most common one, was called magnetite, namely the stone from Magnesia, the city in Asia Minor where the stones were found. It was later observed, first in China and later in Europe, that a magnetic needle floating on water supported by a piece of wood, or revolving around a vertical pivot, turns toward a definite direction relative to earth. We can say that both magnets and the earth generate a magnetic field. The systematic study of magnetic phenomena started in the XVII century, but only in 1820 did Hans Christian Ørsted discover the fundamental fact that electric currents produce a magnetic field, similar to what magnets do. A few years later, André-Marie Ampère, based on a series of experiments, introduced the hypothesis that the only sources of magnetism are electric currents. According to this hypothesis, which was successfully tested through further experiments, a magnetic quantity analogous to the electric charge does not exist in nature. There is no magnetic charge. The magnetism of magnetic materials is due to microscopic currents inside them, at the molecular level.

We shall start by defining the magnetic field \mathbf{B} , with a procedure similar to that which we used for the electric field, on the basis of the force acting on an exploring charge. We shall see that the magnetic force is more complicated than the electric one. Indeed, the magnetic force is found to be proportional to the velocity of the exploring charge and perpendicular to it. In particular, the magnetic force does not act on a charge at rest. We shall see that, as a consequence, every element of a

circuit immersed in a magnetic field is subject to a force. In Sect. 6.3, we shall study the Hall effect, which is a direct consequence of the magnetic force, and which can be exploited to build instruments to measure the magnetic field.

In Sect. 6.4, we shall consider a few examples of motion of charged particles in magnetic fields. We shall see that the magnetic force acting on a point charge changes the direction of its momentum, but, being perpendicular to the motion, cannot change its kinetic energy. In Sect. 6.5, we shall study another important instrument, the galvanometer, used to measure current intensity. It is based on the magnetic force acting on segments of a circuit in a magnetic field.

In the second part of the chapter, starting with Sect. 6.6, we shall study how steady macroscopic currents produce the magnetic field. We shall see that Ampère's law states that the circulation integral of the magnetic field about any closed curve is equal to the total current linked to the curve times a constant, called a vacuum permeability, which is the fundamental constant of magnetism. This property, together with the non-existence of magnetic charges, can be used to specify the curl and the divergence of a magnetic field (under stationary conditions).

In Sect. 6.7, we shall see how to calculate the magnetic field in particularly simple current configurations using the Ampère circulation law alone, profiting from the symmetry of the problem. We shall study the discontinuities of the vector field \mathbf{B} crossing a current sheet. We shall see that its normal component is continuous, the tangent one discontinuous.

Differently from the electrostatic field, the magnetic field is not irrotational, namely it is not conservative. Consequently, the equivalent of the electrostatic potential does not exist. We shall define, in this case, a vector potential, which is such that its curl is equal to the magnetic field. In Sect. 6.8, we shall find the differential equation of the vector potential and see this equation to be exactly the same as the equation of the electrostatic potential in electrostatics, provided we substitute the current density in place of the charge density. This property will allow us to find the vector potential for a few distributions of currents. Once the vector potential is known, we can calculate the magnetic field. In Sect. 6.9, we give some examples of calculation of the vector potential. In Sect. 6.10, we shall give the expression of the magnetic field produced by a given set of currents, the Ampère-Laplace law, and in Sect. 6.11, some examples of direct calculations of magnetic field.

In Sect. 6.12, we shall find the expression of the force between two straight parallel wires carrying steady currents. The definition of the unit of the base quantity in the SI, namely the ampere, is based on the measurement of this force.

In Sect. 6.13, we study the magnetic effects of an important circuit, which is the magnetic dipole. The behavior of the magnetic dipole is similar in several aspects to that of the electric dipole, while it is different in others. In particular, the shapes of the magnetic field of the former and the electric field of the latter are equal at distances from the dipoles that are large compared to their dimensions, while they are very different near the dipoles.

In the last two sections, we discuss the transformation properties of magnetic quantities under changes from one inertial frame to another, namely under rotation

and inversion of the axes in Sect. 6.14, and between two reference frames in relative uniform motion in Sect. 6.15. Going into more detail, in Sect. 6.14, we see that there are quantities that remain invariant under rotations and inversions, called scalars, and those that do not change under rotations but change sign under inversion, called pseudoscalars. Other quantities, for example, the electric field, have three components that change both under rotations and inversions with the same transformation properties as the coordinates. These are the vectors (or proper vectors, if we must be more precise). Still other quantities, such as the magnetic field, behave like the coordinates under rotations but with the opposite sign under inversions. They are called pseudovectors or axial vectors. We shall then discuss the right-hand rule, which seems to indicate an asymmetry under inversion of the axes, and see that this is not the case. In other words, the electromagnetic interaction is invariant not only under rotations but also under inversion of the axes. We state here that the former invariance is a universal law of Nature, while the latter is not. Of the four fundamental interactions, the electromagnetic, the gravitational and the nuclear strong interactions are invariant under inversions, while the weak nuclear interaction is not.

In Sect. 6.16, we shall deal with the following issue. The magnetic force acting on a charged particle at a point in which there is a magnetic field and no electric field depends on the velocity of the particle. In the rest frame of the particle, the magnetic force is zero. However, there is no violation of the relativity principle, because in the latter frame, an electric field exists and the force acting on the particle is electric. The relativity principle is satisfied provided the transformations between frames are Lorentz transformations. We shall see, in particular, that the current density and the charge density, taken together and multiplied by suitable constants, make up a four-vector.

6.1 Preliminary Observations

Ancient civilizations, in particular, the Greeks and the Chinese, long ago noticed the property of certain ferrous minerals to attract pieces of iron. Lodestone, the most common one, which we now know to be an iron oxide ($\text{FeO} \cdot \text{Fe}_2\text{O}_3$), was called magnetite, namely *magnetes lithos*, in Greek, meaning *stone from Magnesia*, the city in Asia Minor where the stones were found. The first written description of the magnetic properties of lodestone is by Thales of Miletus (circa 624–circa 526 BC). The Chinese written reports are from two centuries later. The origins of the compass are unknown. The first written descriptions date to the XIIth century.

The scientific study of magnetism started in the XVIth century, in particular, with the work of William Gilbert (UK, 1544–1603), who published “*De magnete, magneticisque corporibus, et de magno magnete tellure*” (On the magnet, magnetic bodies, and on the great magnet of the earth) in 1600.

It was observed that the magnetic effects of magnets of elongated shapes, in particular of needles, were localized at the two extremes. We can easily observe this

property by dipping a small magnetic bar in iron filings. When we take it out, we see that the filings are attached to the extremes, forming two tufts there, while the central part is clean. The extremes of the magnets are called *magnetic poles*. If we arrange a magnetic bar on a floating support and lay it on water, we observe it spontaneously orienting in the South-North direction. These observations were at the origin of the compass. The original contribution of Gilbert was the hypothesis of earth being a big spherical magnet, for which the needle points to the North pole, rather than to the Polar star, as was previously believed (hence, stars had nothing to do with the effect). The pole of the magnet aiming geographically North is called the north pole (N), the other one, south pole (S).

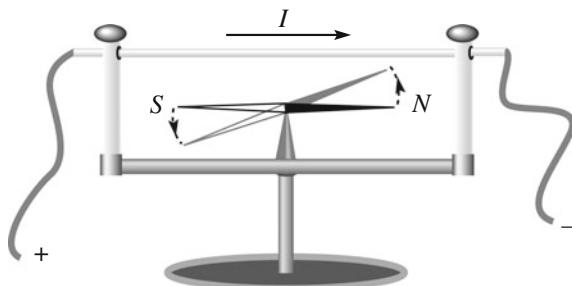
It was very soon found that two equal poles (N and N or S and S) repel, while two different ones (S and N) attract one another. Charles Augustin de Coulomb quantitatively studied the force between the magnetic poles of two different long magnetic bars between 1785 and 1791. He used the same method as in his measurement of the electrostatic force that we saw in Sect. 1.2. He found the force to be similar to the electrostatic one, in particular being inversely proportional to the square distance between the two “magnetic charges”. A theory of magnetostatic phenomena was developed in analogy to the electrostatic theory.

It took almost a century before it was discovered that the theory was wrong. Analogy and symmetry between different situations in physics are often useful guides to our discovering of the physics laws, but they may be sometimes misleading as well. As a matter of fact, no magnetic charge exists in nature, as far as we know. The north and south poles are always coupled, it being impossible to separate one from the other. This can be verified with the simple broken magnet experiment, which was historically first reported in 1269 by Pierre Pelerin de Maricourt (France, XIIIth century). You dip a magnetized bar, which should be quite long and narrow, in iron filings and extract it. The two tufts appear at the extremes. You now cut the bar in the middle and dip both parts in the filings. The observation is that both of them have two filing tufts at the extremes. Checking with another magnet, one can verify that both pieces have a north and a south pole. If you cut each piece in two, you obtain four bars, each with a north and south pole. All the poles have the same strength.

This is a very simple experiment. Much more refined and sensitive experiments have been performed up to the present time, searching for single magnetic poles, called magnetic monopoles, in different types of rocks, including the stones brought from the moon by astronauts, in cosmic rays or as possibly produced by collisions of high energy particle beams from accelerators. All these searches systematically failed. Magnetic monopoles do not exist, as far as we know. The simplest object is the dipole, namely a north and a south pole of the same strength at a certain distance between them. We note here that the word ‘dipole’ in this case is misleading, because the magnetic dipole is not made of two magnetic charges. As we shall soon learn, a magnetic dipole is a loop of current.

In 1820, Hans Christian Ørsted (Denmark, 1777–1851) discovered that an electric current generates a magnetic field. The experiment is shown schematically in Fig. 6.1. He placed a compass on a table. The needle turned with its north pole (black in the figure) toward the geographic North. He placed a wire connected to a

Fig. 6.1 The Ørsted experiment



battery parallel to the needle at a few centimeters above it. When he closed the switch to have the current in the wire, he observed the needle turning to a new equilibrium position. The direction of the deflection is given by the right hand rule. If one puts the right hand near the wire with the palm facing the wire and the direction and sense of the current being parallel with the portion of the arm from wrist to fingers, the thumb points in the direction of the deflection of the north pole of the needle.

At this point, it was clear that not only do magnets and earth produce a magnetic field, but electrical currents do as well. The magnetic effects of currents and the magnetic interactions between currents were accurately studied in the subsequent years, in particular by André-Marie Ampère. We shall come back to his discoveries later in this chapter. We only mention here the fundamental intuition of Ampère, who advanced the hypothesis that electric current is the *unique* source of a magnetic field. He imagined that microscopic currents should exist inside the magnets. These molecular currents should be invisible, but able to produce the field of the permanent magnets according to the same laws as the macroscopic currents that we can control. The magnetic monopole notion started to appear to be a mathematical artifact, deprived of physical meaning. We shall now abandon the historical perspective and start with the study of the magnetic field produced in a vacuum by macroscopic currents. As a matter of fact, these are the simplest phenomena to describe. We shall come back to the more complex study of the magnetism in matter, permanent magnets included, in Chap. 8.

6.2 Magnetic Field

In Sect. 1.3, we operationally defined the electric field under static conditions. The measurement operation consists of measuring the force acting on an electric charge, small enough not to perturb the field, and dividing the result by the charge value. Measuring a magnetic field is more complicated, because the force due to the field is only different from zero on charges in motion, being null for charges at rest. We can measure the magnetic field at a point by studying the trajectories of charged particles of known velocities passing at the point.

Let us consider the charge q going through the considered point several times, with velocities differing in magnitude and direction, and let us measure its acceleration each time. We observe that the directions of the accelerations are, in any case, perpendicular to the velocity, which is different in the different cases, and to a fixed direction in space, which is always the same. We also observe that, for a fixed direction, the force is proportional to the velocity. All the observed properties of the magnetic force can be summarized by defining a vector field $\mathbf{B}(x, y, z)$, which does not depend on time (under the stationary conditions we are considering) and is such that the force acting on the charge q passing at the point (x, y, z) with velocity \mathbf{v} is

$$\mathbf{F}(x, y, z) = q\mathbf{v}(x, y, z) \times \mathbf{B}(x, y, z). \quad (6.1)$$

\mathbf{B} is the magnetic field. Its physical dimensions are those of a force divided by a charge and a velocity, namely $[B] = [FQ^{-1}V^{-1}] = [MT^{-1}Q^{-1}]$. The measurement unit is called a tesla (T), named after Nikola Tesla (Serbia, 1853–1943). The non-SI unit gauss = 10^{-4} T is often encountered in the literature.

We immediately observe that, being that the magnetic field is always perpendicular to the velocity, its work on a charged body is always zero. Indeed, we have

$$\Delta E_K = W = \int_1^2 \mathbf{F} \cdot d\mathbf{l} = \int_1^2 q\mathbf{v} \times \mathbf{B} \cdot \mathbf{v} dt = 0. \quad (6.2)$$

Namely, the magnetic force does not produce changes in the kinetic energy of the charge on which it is acting, or in the magnitude of its velocity, or in the magnitude of its momentum. The force does produce changes in the direction of the velocity and in the direction of momentum. It deflects the charged particles in motion.

Experiments have shown that the superposition principle holds for the magnetic field. This means that if a system of currents 1 produces the magnetic field \mathbf{B}_1 at the point P and another system of currents 2 separately produces the field \mathbf{B}_2 at that point, when the two current systems 1 and 2 are contemporarily present, the field in P is $\mathbf{B}_1 + \mathbf{B}_2$.

In addition, if a charge q moves in a region of space in which both a magnetic field \mathbf{B} and an electric field \mathbf{E} are present, the total force on the charge is the sum of the magnetic and the electric forces, according, once more, to the superposition principle. The force on the point-like charge q is

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \quad (6.3)$$

This expression is completely general. It holds not only under the stationary conditions we are considering but also when the fields \mathbf{E} and \mathbf{B} depend not only on the coordinates but on time as well. The force is called the Lorentz force, named after Hendrik Antoon Lorentz (The Netherland, 1853–1928).

One might observe that, as the Lorentz force is zero when the velocity is zero, it looks as if it contradicts the relativity principle, being different in two inertial frames in relative uniform motion. This is not the case, however. In Sect. 6.16, we shall see how both the electrical and the magnetic terms transform under a Lorentz transformation in such a way that the result is relativistically covariant.

Electrical currents consist of charges in motion. Consequently, a current carrying wire in a magnetic field is subject to magnetic forces. Let us find the expression. We start by recalling that the current density is $\mathbf{j} = n_p q \langle \mathbf{v} \rangle$, where q is the charge of the carriers, $\langle \mathbf{v} \rangle$ is their average velocity and n_p their numeric density. If we consider a wire, of section S small enough to have the same current density at all its points, the current intensity is $I = jS$. Let us consider an infinitesimal element of length ds of such a wire immersed in the magnetic field \mathbf{B} and find the force acting on the element. The magnetic force acting on each carrier, of charge q and velocity \mathbf{v}_q , is $\mathbf{F}_q = q\mathbf{v}_q \times \mathbf{B}$. The total force is obtained by adding all the charges in the volume $S dl$, namely

$$d\mathbf{F} = \sum_{Sdl} q\mathbf{v}_q \times \mathbf{B} = q \left(\sum_{Sdl} \mathbf{v}_q \right) \times \mathbf{B}.$$

The sum over the velocities on the right-hand side is just the average velocity $\langle \mathbf{v} \rangle$ of the carriers times their number, which is $n_p S ds$. Hence, we have

$$d\mathbf{F} = n_p S ds q \langle \mathbf{v} \rangle \times \mathbf{B} = \mathbf{j} S \times \mathbf{B} ds = I ds \times \mathbf{B}$$

where we have profited off the fact that \mathbf{j} and ds have the same positive direction. In conclusion, the force on the circuit element ds is

$$d\mathbf{F} = I ds \times \mathbf{B}. \quad (6.4)$$

Clearly, we cannot observe the force on an element of infinitesimal length. We can only do that on a short element of the circuit. We hang a current-carrying wire, free to move near an electromagnet. When we switch on the magnet, we see the wire moving perpendicularly, both to its own direction and to that of the magnetic field. Performing suitable measurements, we can check that Eq. (6.4) indeed corresponds to the Lorentz force acting on the charges moving through the wire. Historically, Eq. (6.4) was discovered before the Lorentz force and is sometimes called the 2nd Laplace law, from Pierre-Simon Laplace (France, 1749–1827).

The total force acting on a finite segment Γ between, say, sections A and B of a circuit carrying the current I immersed in the magnetic field \mathbf{B} is obtained by integration

$$\mathbf{F} = I \int_{\Gamma,A}^B ds \times \mathbf{B}. \quad (6.5)$$

6.3 Hall Effect

The method for measuring a magnetic field that we have described in the previous section is conceptual, but cannot obviously be used in practice. A conceptually simple but also practical method is based on a physical effect discovered in 1879 by Edwin Hall (USA, 1855–1938) and is called the Hall effect.

Consider a conductor shaped like a parallelepiped with a rectangular base of sides a and b and a small height c . It is immersed in the magnetic field \mathbf{B} perpendicular to its base and practically uniform in the size of the conductor, as shown in Fig. 6.2. We have soldered two wires (not shown in the figure) to the central points of the two opposite faces ac and have injected the constant current I . We have soldered another couple of wires to the centers of the faces bc and connected them to a voltmeter to measure the potential difference between those points. We do observe a potential difference, called the Hall voltage, which we indicate with V_H . This is the Hall effect. It is found that V_H is proportional both to the current intensity I and to the magnetic field B .

Figure 6.2a explains the effect in the case of positive carriers, Fig. 6.2b in the case of negative carriers. One sees that the magnetic force $\mathbf{F}_m = q\mathbf{v} \times \mathbf{B}$ has the same direction in both cases. Its intensity, being that the field and velocity are mutually perpendicular by construction, is $F_m = qvB$.

Initially, the carriers under the action of the magnetic force move to the left in the figure (in both cases). This causes the development of a charge profile decreasing from left to right, which soon repels newly arriving carriers. At equilibrium, the corresponding electric force F_e is equal and the opposite of the magnetic force, namely it is $F_e = F_m$. Calling $E = F_e/q$ the corresponding electric field, we have $E = |\langle \mathbf{v} \rangle|B$.

This is the electric field which appears as a potential difference between the two sides of the conductor, namely as $V_H = aE = |\langle \mathbf{v} \rangle|aB$.

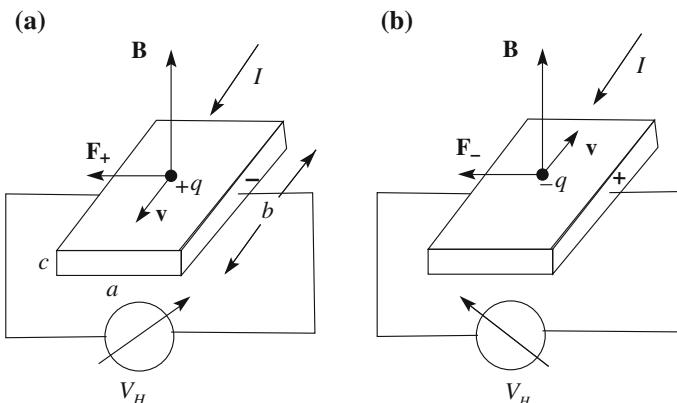


Fig. 6.2 Hall effect. **a** Positive carriers, **b** negative carriers

We can express $\langle \mathbf{v} \rangle$ in terms of the current density, and then of the current intensity. We shall limit ourselves to the cases in which carriers of only one sign are present. Thus, we have $I = jac = n_p q |\langle \mathbf{v} \rangle| ac$, where n_p is the carrier number density. Hence, finally,

$$V_H = \frac{IB}{n_p q c}. \quad (6.6)$$

Let us now comment on the result. Firstly, Eq. (6.6) explains the observation of the Hall voltage as being proportional both to the current intensity and to the magnetic field magnitude. Secondly, the Hall voltage is inversely proportional to the charge of the carrier. Consequently, the voltage *sign* depends on the sign of the charge of the carrier, which can so be determined.

We can enhance the effect by decreasing the thickness c of the conductor and by working with lower carrier densities. This is why the effect is easier to detect in semiconductors than it is in metals. Consider, for example, copper, in which, as we have seen in Sect. 5.1, $n_p = 1.7 \times 10^{29} \text{ m}^{-3}$, and let us use a quite thin conductor with $c = 0.1 \text{ mm}$. Equation (6.6) shows that the Hall voltage, even for an intense field on the order of one tesla and current intensities on the order of one ampere, is only of a few microvolts. In a semiconductor, the carrier density may easily be 10^{-10} times smaller than in a metal. If we use current intensities of a few milliampere, which are small enough not to damage the semiconductor, we obtain easily measurable Hall voltages in the millivolt range, even with magnetic fields of 10^{-4} T .

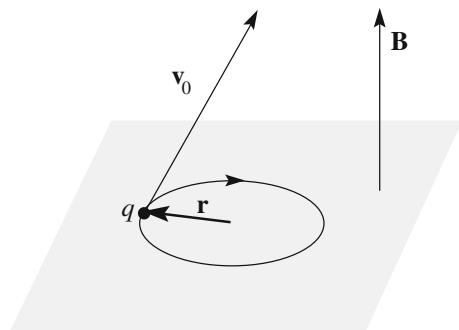
QUESTION Q 6.1. A current of 1.0 A flows in a rectangular slab of silver has a thickness of 1 mm . The Hall voltage measured in a magnetic field of 1 T perpendicular to the slab is $0.1 \mu\text{A}$. What is the carrier density? \square

Hall probes based on these principles are cheap, commercially available, and easy to use. Note, however, that the proportionality constant $n_p q c$ in Eq. (6.6) is, in general, not accurately known. Consequently, the Hall probes give relative measures of the field. To have an absolute measurement, they must be calibrated in a known magnetic field. Note also that the carrier number density of a semiconductor is a strong function of temperature. Consequently, the calibration must be done at the same temperature as the subsequent use.

6.4 Motion of Charges in a Magnetic Field

In this section, we study the motion of a point charge in a stationary magnetic field \mathbf{B} . Let us start with the simplest situation, in which the field is uniform and the initial velocity \mathbf{v}_0 is in a plane perpendicular to \mathbf{B} , as in Fig. 6.3. Let q be the charge and m the mass of the particle.

Fig. 6.3 The trajectory of a point charge moving in a plane perpendicular to a uniform magnetic field



The magnetic force on the particle at the initial instant is

$$\mathbf{F} = q\mathbf{v}_0 \times \mathbf{B}. \quad (6.7)$$

The direction of the force belongs to the plane normal to \mathbf{B} containing \mathbf{v}_0 . Being that the force is always perpendicular to the motion, the velocity is constant in magnitude. The motion is circular and uniform. Let \mathbf{r} be the position vector taken from the center of the orbit. The known relation between linear and angular velocity is $\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$. Being that $\boldsymbol{\omega}$ is constant, the (centripetal) acceleration is $\mathbf{a} = \boldsymbol{\omega} \times \mathbf{v}$. The acceleration is equal to the force, which is the magnetic one, divided by the mass, namely $\boldsymbol{\omega} \times \mathbf{v} = (q/m)\mathbf{v} \times \mathbf{B}$, which finally gives us

$$\boldsymbol{\omega} = -(q/m)\mathbf{B}. \quad (6.8)$$

Note that the angular velocity is independent of the initial velocity. This is true for the period T and the frequency $v = 1/T$ as well. These quantities are

$$T = \frac{2\pi}{\omega} = \frac{2\pi m}{B q}; \quad v = \frac{B q}{2\pi m}$$

The orbit radius depends on the velocity. Being that $r = v_0/\omega = v/\omega$, we have

$$r = mv_0/(qB) = mv/(qB). \quad (6.9)$$

Let us now look at a few interesting examples of this motion.

The cyclotron is a charged particle accelerator invented by Ernest Orlando Lawrence (USA, 1901–1958) in 1932. Figure 6.4a is a reproduction of the scheme submitted with the patent request, which was granted in 1934. Accelerators employing high voltages, like the van der Graaf accelerator, are linear structures, whose dimensions increase with the accelerating voltage. The cyclotron works without high voltage and using a circular structure. This allows for reaching much higher energies in compact configurations. The first cyclotron was made out of brass and was only 4 in. (10 cm) in diameter. It could literally be held in one hand.

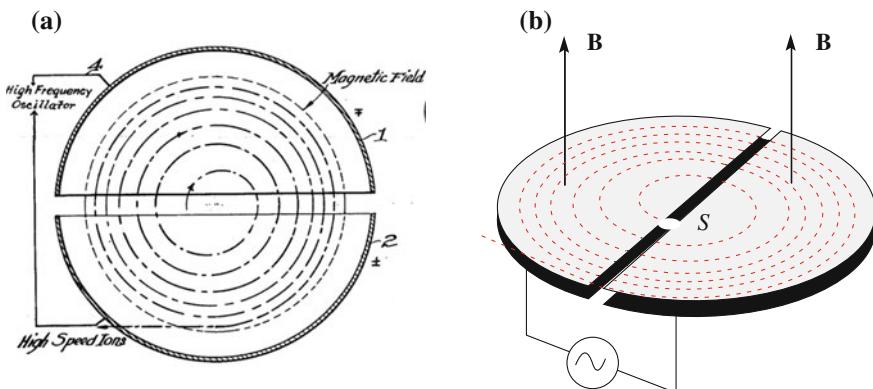


Fig. 6.4 The cyclotron. **a** Original scheme in the Lawrence patent request; **b** scheme of the working principle

In the following years, cyclotrons of increasing size were built, which were able to accelerate ions to increasing energies. A 152 cm diameter cyclotron built in 1939 could accelerate ions to 16 MeV.

Figure 6.4b helps us to understand the working principle, which is based on the independence of the angular velocity of the orbit radius, a property that we have just seen. An important component of the device is the electromagnet, which has two planar poles facing one another. The magnet produces a uniform magnetic field in the space between the poles, where the cyclotron is installed. The cyclotron itself, which is under a vacuum, contains two hollow electrodes called Dees, for their shape, in the plane perpendicular to the field. The source of ions, S in the figure, is located in the gap between the Dees near the center. Between the electrodes, an emf is applied that produces an electric field, which accelerates the ions initially to the left in the figure. The ions enter the Dee and become shielded from the electric field, having only the magnetic force acting on them. Consequently, they move along a semicircle.

If the emf were constant in time, the ions, when returning to the gap between the Dees in the opposite direction, would meet a decelerating electric field. The trick is to have an emf periodically varying in time with half a period exactly equal to the time taken by the ions to describe half a circle. In such a way, when the ions exit the gap and feel the electric field, they find it inverted and are accelerated again. The process repeats itself in the Dee on the right, and then on the left, and so on. If ΔV is the value of the emf and q is the charge of the ion, the kinetic energy gained at each passage in the accelerating gap is $\Delta U_K = q\Delta V$. As the velocity increases, the orbit radius increases as well, according to Eq. (6.9), but the time taken to move along half a circle remains constant. Hence, a emf generator of fixed frequency can be used. Otherwise, it would be impractical.

The maximum energy is reached when the orbit radius R is equal to the radius of the cyclotron. This radius is limited by the diameter of the poles of the magnet, which is the most expensive part of the system. The maximum energy is

$$U_K = \frac{1}{2}mv^2 = \frac{q^2B^2R^2}{2m}. \quad (6.10)$$

A theoretical limit to the achievable acceleration with a cyclotron originates through the fact that the equations of motion we wrote are non-relativistic, namely they hold for velocities that are small compared to the speed of light. At higher velocities, the period is no longer independent of velocity and the cyclotron, in this simple configuration, ceases to work.

Note the different functions of electric and magnetic fields. The electric field accelerates the ions, the magnetic field changes the direction of their momentum, guiding them on the trajectory. The particle accelerators at relativistic energies, up to the highest ones, employ a number of different technologies, but the functions of the electric and magnetic fields are those we have just mentioned in any case.

Crossed electric and magnetic fields are used to build velocity selectors for charged particles. The charged particle sources, like a hot metal wire for electrons or an ampule containing a gas in which an electric discharge takes place for ions, produce particles over a range of velocities. If we want to select a beam of particles in a narrow velocity range, we can use the selector shown in Fig. 6.5.

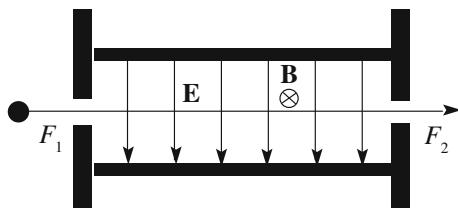
A parallel plate capacitor is used to produce a uniform electric field \mathbf{E} , and an electromagnet to produce a uniform magnetic field \mathbf{B} perpendicular to \mathbf{E} . The electric force $q\mathbf{E}$ and magnetic force are parallel and, choosing the senses of the fields properly, opposite. Two slits F_1 and F_2 select the electrons traveling along a straight trajectory. The electron velocity must be such that the resultant force is zero, namely that $qE = qvB$. The selected velocity, within an interval dependent on the diameters of the slits and their distance, is

$$v = E/B. \quad (6.11)$$

The velocity can be chosen adjusting one of the two fields.

QUESTION Q 6.2. We adjust the velocity selector to have electrons of a certain velocity throughout. If we now input protons with the same velocity, will they go through? \square

Fig. 6.5 Charged particles velocity selector. The symbol \otimes , representing the fletching of a vector arrow seen from the back, means that \mathbf{B} is directed inside the drawing



A mass spectrograph is an instrument capable of separating different isotopes. The name is analogous with optics. An optical spectrograph, for example, a prism, separates a beam of white light into its components of different colors, which have different frequencies. A light spectrum is a diagram of the intensity as a function of frequency. Similarly, an isotope mass spectrum is a diagram showing the isotopic abundance as a function of mass, or, to be more precise, of the ratio m/q . A high-resolution mass spectrometer was invented by Arthur Dempster (Canada-USA, 1886–1950) in 1918. The device is shown schematically in Fig. 6.6.

The entire apparatus is under a vacuum. A source S , located in a small chamber, produces the ions of the element to be studied. Ions of different charge, q , and mass, m , are present. Two electrically insulated slits F_1 and F_2 select a beam of ions that are accelerated by a potential difference, typically of $V = 50\text{--}100\text{ V}$, applied between the slits. The kinetic energy of an ion, which is negligible at the source, is

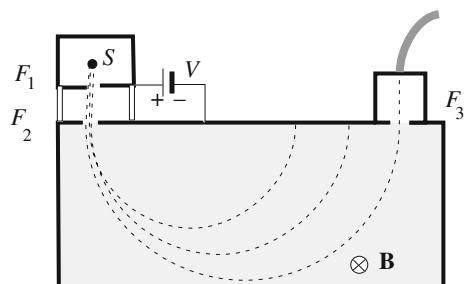
$$\frac{1}{2}mv^2 = qV \quad (6.12)$$

when it crosses F_2 . After F_2 , the ions enter a larger chamber (typically 10–15 cm across) that is immersed in a uniform magnetic field \mathbf{B} , directed inside the drawing in Fig. 6.6. The trajectories of the ions in the main chamber are semi-circles of radius $R = mv/(qB)$, which, with v given by Eq. (6.12), is

$$R = \frac{1}{B} \sqrt{2(m/q)V}. \quad (6.13)$$

We see that ions of different mass to charge ratio m/q travel along orbits of different radii. In practice, all the ions are singly ionized and have the same charge. Figure 6.6 gives us an example with three different masses. A third slit F_3 is located on the right-hand side of the chamber, in order to select only those ions whose orbit diameter is equal to the distance D between F_2 and F_3 . The ions getting through F_3 enter a third chamber in which their current is measured by a sensitive electrometer. The distance D is fixed, and we obtain a mass spectrum by varying the magnetic field and measuring the ion current as a function of the field. We obtain a

Fig. 6.6 The Dempster mass spectrograph



number of peaks, each corresponding to a different mass of height proportional to the abundance.

An important property of every spectrograph, the Dempster mass spectrograph included, is its resolving power. Consider having two ion species in the beam with masses m and $m + \Delta m$. If Δm is too small, both components may be able to enter F_3 , for the same values of V and B . We say that the components are not resolved. As a matter of fact, in order to enter F_3 , the orbit diameter does not need to be exactly equal to the distance D between the centers of F_2 and F_3 , because the slits have a certain width. Assuming, for simplicity, that both have the same width ΔD , the condition is $D - \Delta D \leq 2R \leq D + \Delta D$. Differentiating Eq. (6.13), we obtain

$$\frac{\Delta m}{m} = \pm 2 \frac{\Delta D}{D}, \quad (6.14)$$

where the factor 2 on the right-hand side comes from m being proportional to R^2 and consequently to D^2 . Clearly, we can increase the resolving power (within certain limits) by decreasing the widths of the slits. Let us look at the orders of magnitude. Let $D = 60$ mm be the distance between the slits and $\Delta D = 0.3$ mm their width. Equation (6.14) gives a resolving power of 1 %. This means that we can separate two isotopes differing in mass by 1 u (unified atomic mass unit, which is approximately one nucleon mass), which is the smallest mass difference between ions, up to 100 u.

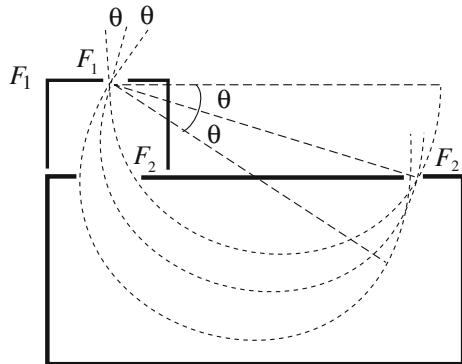
However, reducing the width of the slits, while increasing the resolving power, reduces the useful ion flux as well. As a matter of fact, no spectrometer can work if it is not capable of focusing. Indeed, having a small source S is not enough; we must also be able to produce an image of S , as small as possible in the location of the detector. This is at F_3 , in our case. An important property of the Dempster spectrograph is its focusing power after a 180° deflection.

Figure 6.7 shows the trajectories of three ions of the same mass and the same charge entering F_2 at different angles, one along the ideal trajectory, one at the angle $+\theta$ and one at $-\theta$ to it. Initially, the trajectories diverge, but after 90°, they converge back to become very close to one another at 180°. At this point, we have a focus and we can locate the detector. In this way, it will detect ions within a certain angular acceptance. A spectrograph of zero acceptance cannot work, because the flux it accepts is null. Let us look at the problem quantitatively.

An ion entering along the ideal trajectory, namely at $\theta = 0$, hits the wall at the distance D from its entering point, as we have seen. An ion entering at the angle θ hits the wall at the distance equal to the chord $c = d \cos \theta$. The distance between these points is $\Delta D = D(1 - \cos \theta) \cong D\theta^2/2$, where we have approximated the cosine as $\cos \theta \cong 1 - \theta^2/2$. We then define as focusing power of the instrument

$$\frac{\Delta D}{D} = \frac{\theta^2}{2}.$$

Fig. 6.7 Focusing property of the Dempster spectrographic

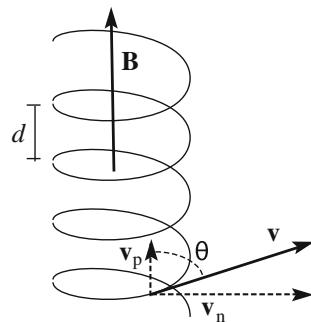


With the data of the above considered example, $\Delta D/D = 5 \times 10^{-3}$, and hence the acceptance is $\theta = 0.1$ rad. Namely, F_2 accepts ions in an angular range of ± 0.1 rad equal to $\pm 5.7^\circ$ to the ideal direction. This is quite a reasonable value to work with. We have just mentioned that Dempster used his spectrograph to separate and measure the relative isotopic abundance of several elements, including Li, Na, K, Mg, Ca and Zn.

One quarter of a century after its discovery, arrays of the Dempster spectrograph on an enormous scale were used by the Manhattan project at Oak Ridge to enrich in ^{235}U natural uranium to the grade necessary for the bomb. As the reader has noticed, the Dempster spectrometer is quite similar to one half of a cyclotron. And it was just Lawrence who modified the Dempster arrangement to increase the ion current while still retaining adequate resolution in mass separation. As part of the necessary R&D program, the 37-in. cyclotron at Berkeley was dismantled on November 24, 1941, and its magnet used to create the first *calutron*. Its name came from a mash-up of “California University” and “cyclotron”. Research led to a “racetrack” configuration of the basic unit. This unit, or building, consisted of two magnets with forty-eight gaps, each gap containing two vacuum tanks, resulting in 96 source and collector pairs. Ten such buildings, with a total of about 2,000 sources and collectors, were necessary to separate 100 g of ^{235}U daily, as requested by the project. The Y-12 plant, as it was called, started production in 1943 at Oak Ridge (Tennessee).

Let us now consider the motion of a point charge still in a uniform magnetic field, but not in a plane perpendicular to the field, as shown in Fig. 6.8. Let us decompose the velocity in two components, one perpendicular and one parallel to \mathbf{B} . We indicate them with \mathbf{v}_n and \mathbf{v}_p , respectively, and we have $\mathbf{v} = \mathbf{v}_n + \mathbf{v}_p$. Let θ be the angle between \mathbf{v} and \mathbf{B} . The motion in the plane normal to \mathbf{B} is a uniform circular motion with velocity magnitude $v_n = v \sin \theta$. The motion in the direction of the field is uniform with velocity $v_p = v \cos \theta$, because the magnetic force component is zero.

Fig. 6.8 Motion of a charge in a uniform magnetic field



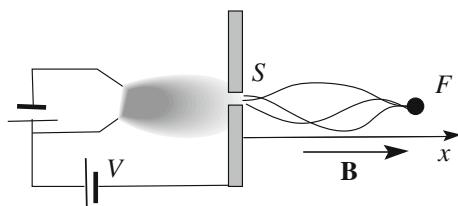
Hence, the trajectory is a cylindrical helix. The period of the circular projection is $T = 2\pi m/(qB)$. The pitch d of the helix is the distance traveled in the motion parallel to the field in a period, namely

$$d = v_p T = 2\pi \frac{m v \cos \theta}{q B}. \quad (6.15)$$

We see that the pitch is independent of the radius. This property is exploited for a second method, different from that of the Dempster, of magnetic focalization of charged particle beams, electrons, for example. Figure 6.9 shows the scheme of the device, which is under a vacuum. A metallic wire heated by an electric current emits electrons. An applied potential difference V accelerates the electrons towards a positive electrode in which there is a small hole S .

Let x be the direction of the beam axis. Initially, the electrons have quite small velocities with different directions. All the electrons that exit the hole have the x -component of their velocity accelerated by the electric field to $v_x = \sqrt{2qV/m}$, which is large compared to the initial velocity. In addition, each electron has small components of its velocity perpendicular to the x -direction. Due to these transversal components, the electron beam widens as it moves away from the hole. Let us try to focalize the beam, namely to have an image of the effective source S at a certain distance. We can do that with uniform magnetic field \mathbf{B} in the x -direction. All the electrons in their helical trajectories will converge back at the point F of the axis at a distance equal to the pitch of the helix, namely

Fig. 6.9 Magnetic focalization with parallel field



$$d = v_p T = \sqrt{\frac{2qV}{m}} \frac{2\pi m}{qB} = \frac{2\sqrt{2}\pi}{B} \sqrt{\frac{mV}{q}}$$

An image of the source S is formed in the focus F at the distance d given by this equation.

Consider now the motion of a point charge in a non-uniform magnetic field. In regions in which the field does not vary too much, the trajectory is still similar to a helix. However, when the particle moves towards regions of higher field, the radius of the helix gets smaller, as suggested by Eq. (6.9). In addition, the pitch of the helix decreases as well. Indeed, Fig. 6.10a shows that where the field lines are not parallel (namely the field is not uniform), the Lorentz force is not in the plane of the particle orbit. Let us consider its two components, one in the plane of the orbit and one perpendicular to the plane. The former gives the centripetal acceleration of the circular motion, the latter is, in any case, directed toward the regions of lower field. This component slows down, so to speak, the translation velocity of the circular orbit if the particle is moving toward a higher intensity field, and accelerates it if moving towards a weaker field.

As a matter of fact, the orbit tends to include the same bunch of field lines as it moves. If the field varies enough, the limit can be reached at which the orbit slows down and stops, as shown in Fig. 6.10b, and after that, moves back towards weaker field regions. This phenomenon is called a magnetic mirror. It can also be shown that when the “motion of the orbit” slows down, the motion of the particle on the orbit is accelerated, in such a way that the sum of the kinetic energies of the two components of the motion remains unaltered. Kinetic energy is conserved under these conditions, because the acting force is the magnetic force, whose work is null. Non-uniform magnetic field structures are used to confine the plasmas in research programs aimed at nuclear fusion for energy production.

A naturally-occurring charged particle confinement phenomenon happens in the Van Allen radiation belts in the ionosphere. Regions of intense radiation surrounding the earth were discovered in 1958 by James Van Allen (USA, 1914–2006) with Geiger-Müller counters on board the NASA satellites Explorer 1 and Explorer 3. The belts contain a plasma of charged particles, electrons and ions, emitted by the

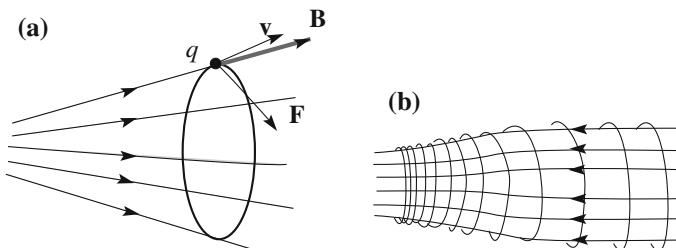


Fig. 6.10 Trajectories in a non-uniform magnetic field

sun, that get trapped in the earth's magnetic field, as shown schematically in Fig. 6.11. The inner belt extends in height typically between 1000 and 6000 km and in celestial latitude between about $+65^\circ$ and -65° from the celestial equator.

The earth's magnetic field is, in a first approximation, a dipole field (which we shall study in Sect. 6.13). This is the field represented in the figure, but we shall come back to a better approximation soon. Note that the North and South poles are inverted relative to a "normal" dipole. This is due to the fact that the north pole of a magnetic dipole, like the compass needle, points to the North pole of earth. Figure 6.11 shows examples of trajectories of charged particles of a certain energy. The pitch of the helix that is large at the equator becomes smaller and smaller towards the poles. With that energy, the charged particle cannot reach the poles, because the magnetic field is too high and the trajectory bunches back. In this way, charged particles get trapped in a belt.

As a matter of fact the magnetic field near earth has two main components, one due to the earth itself, which is mainly a dipole, and one due to the *solar wind*, which is the flux of charged nuclei coming from the sun. The resulting field is shown in Fig. 6.12.

The field is flatter on the side of the sun, namely the field lines are closer to earth than they would have been if there were no solar wind, and has a long tail on the opposite side. On the other hand, the charged particles, when meeting the relatively higher field zones near earth, are deflected by the Lorentz force and do not reach the earth's surface. The magnetic field protects the surface from the intense particle radiation coming from sun. As a matter of fact, the radiation penetrates deeper into

Fig. 6.11 Scheme of a particle trajectory in the Van Allen radiation belts

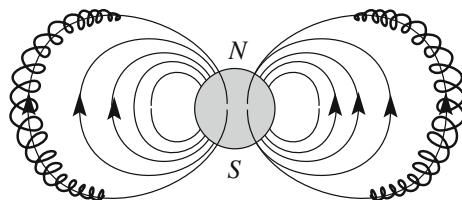
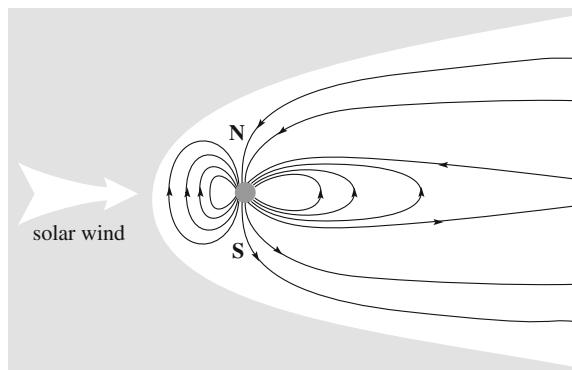


Fig. 6.12 The solar wind and the magnetosphere



the atmosphere at high latitudes, where the velocities of the charges are almost parallel to the field and consequently the Lorentz force is small.

The solar wind and the magnetic field it originates are not constant in time at all, but vary with solar activity. The number of solar spots, which are a mark of solar activity, varies over a period of 11 years. Particularly violent phenomena of mass emission in the solar corona may happen at any time, but their frequency is correlated with the solar spot period. When the particle shower reaches and interacts with the earth's magnetic field, it can perturb it considerably. During such magnetic storms, the phenomenon of the aurorae, taking place at high latitudes, is particularly spectacular. The most intense magnetic storms can severely perturb radio communications, navigation systems and even the electric power distribution grids.

6.5 Galvanometer

Galvanometers are sensitive instruments for measuring current intensity. The instrument is named after Luigi Galvani (Italy, 1737–1798), who discovered in 1791 that electric current would make the legs of a dead frog jerk, but did not develop the instrument.

In Sect. 6.2, we saw that a section of a circuit AB carrying the current of intensity I immersed in a magnetic field \mathbf{B} is subject to the force

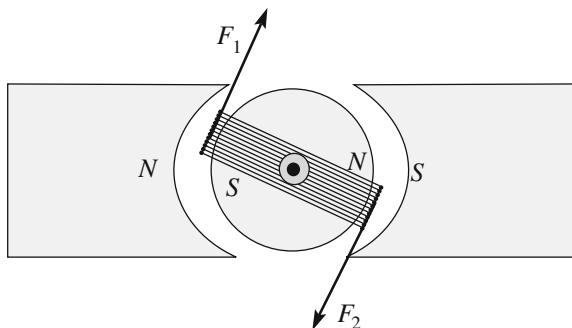
$$\mathbf{F} = I \int\limits_{\Gamma, A}^B d\mathbf{s} \times \mathbf{B}. \quad (6.16)$$

The galvanometer exploits this property to measure I .

Figure 6.13 shows schematically a moving coil galvanometer seen from above. The poles of a small permanent magnet are shaped with concave cylindrical surfaces, as shown in the figure. A cylindrical iron core is located on the axis of the gap. This configuration guarantees a radial magnetic field in the remaining air gaps on the two sides of the core. The intensity of the field is independent of the angle. A rectangular wire coil carrying the current to measure is fixed to a light mechanical support, which is the arm of a sensitive torsion balance suspended to a thin metal wire. The figure shows the suspension point as a black dot. The area of the coil is S and its vertical sides that are immersed in the magnetic field have length a . When the coil rotates about its suspension wire its vertical sides move in the gaps between the poles and the core, in which they find a field always perpendicular to them.

If \mathbf{a} is one of the vertical sides, considered as a vector in the direction of the current, the magnetic force on it is $\mathbf{F}_1 = I\mathbf{a} \times \mathbf{B}$. Its magnitude is $F_1 = IaB$ and its direction is perpendicular to the side. The force \mathbf{F}_2 on the other vertical side $-\mathbf{a}$ is equal and opposite. Hence, the two forces are a couple, with moment of magnitude

Fig. 6.13 Schematic view of a moving coil galvanometer seen from above



$\tau_I = IBab = IBS$. In order to increase the sensitivity, the coil has a number of turns. If this is N the total moment on the coil is

$$\tau_I = NIBS. \quad (6.17)$$

The torsion moment τ_M of the suspension wire is proportional to the rotation angle α , namely

$$\tau_M = -K\alpha, \quad (6.18)$$

where K is the torsional elastic constant. The equilibrium is reached at the angle, call it α^* , at which $\tau_I + \tau_M = 0$, namely for

$$\alpha^* = IBSN/K. \quad (6.19)$$

Measuring α^* , we obtain I once we know the other quantities. In practice, to obtain accurate current measurements, one needs to calibrate the instrument by measuring known current intensities.

QUESTION Q 6.3. The rectangular coil of a galvanometer similar to that of Fig. 6.13 has an area $S = 10^{-4} \text{ m}^2$ and $N = 80$ turns. The magnetic field is $B = 10^{-2} \text{ T}$ independently of the orientation of the coil. What is the torsional elastic constant K of the suspension wire if a current of 0.2 mA produces an angular deflection of 10° ? \square

6.6 The Magnetic Field of Steady Currents

We shall now start to study the relation between electric currents and the magnetic field they produce, in time independent conditions. In general, a vector field, such as the magnetic field, is known once its sources and its curls are known. We must find two expressions, one for $\nabla \cdot \mathbf{B}$ and one for $\nabla \times \mathbf{B}$.

As we already stated, magnetic charges do not exist in nature. The simplest sources of a field are point-like, such as the electric charges for the electric field, which are points from which lines of force of the field originate (sources), or in which they enter (sinks). The magnetic field does not have any source or sink. We can express this property in terms of the flux of a magnetic field through an arbitrary oriented surface Σ , namely in terms of

$$\Phi_B = \int_{\Sigma} \mathbf{B} \cdot \mathbf{n} d\Sigma,$$

where \mathbf{n} is the positive unit vector normal to Σ . The physical dimensions of Φ_B are of a magnetic field times an area. Its unit is called a weber, $1 \text{ Wb} = 1 \text{ T} \times 1 \text{ m}^2$, after Wilhelm Weber (Germany, 1804–1891). Now, the flux of \mathbf{B} through a closed surface Σ is proportional to the field lines crossing Σ , which is zero. Namely, the flux of \mathbf{B} through any closed surface is zero:

$$\oint_{\Sigma} \mathbf{B} \cdot \mathbf{n} d\Sigma = 0. \quad (6.20)$$

The same property can be expressed in differential form, exactly as we did for the electric field, using the Gauss divergence theorem. If V is the volume enclosed by Σ , we have

$$\int_V \nabla \cdot \mathbf{B} dV = \oint_{\Sigma} \mathbf{B} \cdot \mathbf{n} d\Sigma = 0.$$

But, as V is arbitrary, we also have

$$\nabla \cdot \mathbf{B} = 0. \quad (6.21)$$

This equation is analogous to $\nabla \cdot \mathbf{E} = \rho/\epsilon_0$ and expresses the non-existence of the magnetic analog of the charge density. The fields having identically zero divergence are said to be solenoidal.

We have seen in Sect. 6.1 how Hans Christian Ørsted discovered that electric currents give origin to magnetic fields. The magnetic actions of electric currents were then studied in a series of experiments by Jean-Baptiste Biot (France, 1777–1862), Félix Savart (France, 1791–1841) and, above all, André-Marie Ampère (France, 1775–1836).

Consider a rectilinear wire and dispose it vertically, going through the hole in a horizontal plastic plate. If we pour iron filings on the plate and send a current through the wire, we shall see how the filings form circles with their center on the wire. The tiny iron needles orient in the direction of the magnetic field in their position. In such a way, they materialize, so to speak, the magnetic field lines. The pattern is sometimes called a magnetic spectrum.

The accurate measurements done by Biot and Savart have shown that the magnetic field at a given point is in the plane normal to the field through that point and is normal to the wire. The field lines are circles with centers on the wire. The direction of the field is given by the right hand rule. This can be expressed by saying that an observer lying along the current with the positive direction entering through his feet and coming out of his head sees the field lines as having a counter-clockwise direction. This is equivalent to the right-hand rule.

The Biot-Savart law states, in addition, that the magnitude of the field of a steady rectilinear current is inversely proportional to the distance r from the current and is proportional to the current intensity I . To find the equation, let us consider one of the oriented circles made by the field lines. Let us call it Γ and let us calculate the circulation integral of \mathbf{B} about Γ . Taking into account that B has the same value at all the points of the line and is always tangent to the line in the positive direction, we have

$$\oint_{\Gamma} \mathbf{B} \cdot d\mathbf{s} = B 2\pi r$$

As B is inversely proportional to r , the integral is independent of r . Being that B is proportional to the current intensity I , the circulation integral is proportional to I as well. The proportionality constant is the fundamental constant of magnetism, called *vacuum permeability* or *permeability of the free space*, and is indicated with μ_0 . We then have

$$\oint_{\Gamma} \mathbf{B} \cdot d\mathbf{s} = \mu_0 I. \quad (6.22)$$

The value of μ_0 in SI is fixed, as we shall see, by the definition of the ampere to

$$\mu_0 = 4\pi 10^{-7} \text{ N A}^{-2}. \quad (6.23)$$

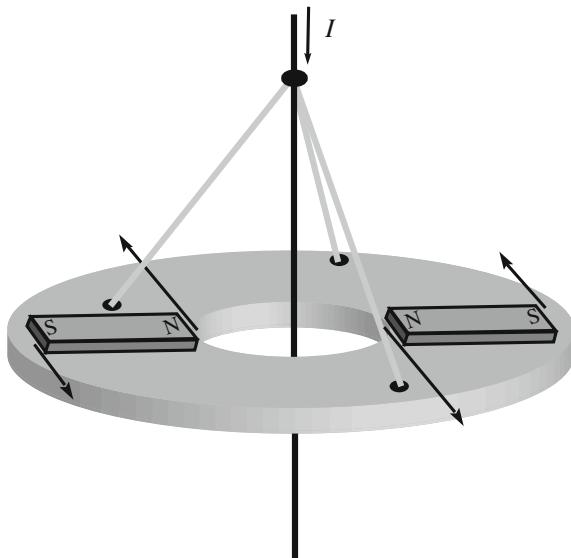
We can now conclude that the magnitude of the magnetic field generated by a continuous rectilinear current is

$$B = \frac{\mu_0 I}{2\pi r}. \quad (6.24)$$

An experiment credited to James Clerk Maxwell (UK, 1831–1879) for demonstrating the $1/r$ dependence of the magnetic field of a rectilinear current is shown in Fig. 6.14. The experiment is designed following the “null” method, which was a mark of Maxwell. In this method, if the law being checked is exactly as expected, two effects balance one another out and nothing happens.

In the arrangement shown in the figure, two torques are present, one on the north poles and one on the south poles of the two magnets, which are equal to one

Fig. 6.14 The Maxwell demonstration of the Biot-Savart law

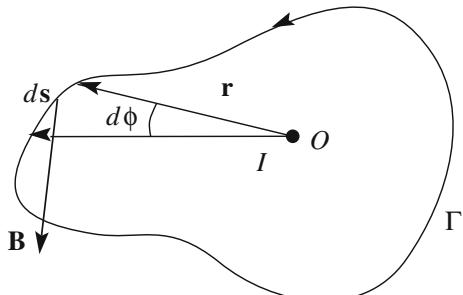


another. The moments of the torques are the products of the forces times the arms. The arm lengths are in the ratio of the distances of the north and south poles from the current, respectively. Only if the forces are in the same ratio inversely do the two effects balance. The force on a magnetic pole being proportional to the magnetic field, the observed effect is null only if the field is proportional to $1/r$ as well.

We shall now prove that Eq. (6.22) holds not only for a circle, but also for every closed curve Γ linked with the current. The positive direction of Γ must be chosen in order for an observer laying on the current with the positive direction of the current entering through his feet and flowing out of his head to see the direction of Γ counter-clockwise. Let us show that first, assuming Γ to be planar and in a plane normal to the current. This is the plane in Fig. 6.15.

Let us think of the infinitesimal segment ds of Γ as being the vector sum of two components, an infinitesimal circular arc and an infinitesimal radial segment. The contribution of the latter to the integral is zero, because it is perpendicular to \mathbf{B} .

Fig. 6.15 Calculating the circulation integral of \mathbf{B} around a planar curve linked with the current



We take the origin O of the reference frame at a point at which the current crosses the plane of Γ . Let \mathbf{r} be the position vector of $d\mathbf{s}$ and $d\phi$ the angle under which $d\mathbf{s}$ is seen from O . The circumference arc we are considering has length $rd\phi$ and the orientation of \mathbf{B} . Hence, we have $\mathbf{B} \cdot d\mathbf{s} = Brd\phi = \frac{\mu_0 I}{2\pi r} rd\phi = \frac{\mu_0 I}{2\pi} d\phi$. We see that this quantity is independent of r . Its integral around Γ is the integral in ϕ over the entire 2π angle, namely

$$\oint_{\Gamma} \mathbf{B} \cdot d\mathbf{s} = \frac{\mu_0 I}{2\pi} \int_0^{2\pi} d\phi = \mu_0 I.$$

Let us consider now a non-planar curve. Every infinitesimal segment $d\mathbf{s}$ of the curve can be considered as being the vector sum of three components, one parallel to the current, one radial and one on a circular arc. The contributions of the first two components are zero, because both are normal to \mathbf{B} , and we are back at the previous case.

Let us now calculate the circulation of \mathbf{B} around a closed curve that does not link the current. We shall limit the discussion, for the sake of simplicity, to a planar curve Γ in a plane normal to the current. We leave to the reader the obvious generalization.

Let O again be the point at which the current cuts the plane of Γ . As one sees in Fig. 6.16, every infinitesimal angle $d\phi$ cuts two infinitesimal segments on Γ (or an even number of such segments). The integrand $\mathbf{B} \cdot d\mathbf{s}$ has, on the two segments, equal and opposite values and the two contributions cancel one another out. It follows that the circulation integral is zero.

In conclusion, we can state that $\oint_{\Gamma} \mathbf{B} \cdot d\mathbf{s} = \mu_0 I$ if I is linked to Γ , $\oint_{\Gamma} \mathbf{B} \cdot d\mathbf{s} = 0$ if I is not linked to Γ .

If several stationary currents exist, like I_1, I_2, I_3 in Fig. 6.17, the superposition principle allows us to conclude that

$$\oint_{\Gamma} \mathbf{B} \cdot d\mathbf{s} = \mu_0 I_{\text{link}}. \quad (6.25)$$

Fig. 6.16 Calculating the circulation integral of \mathbf{B} around a planar curve not linked with the current

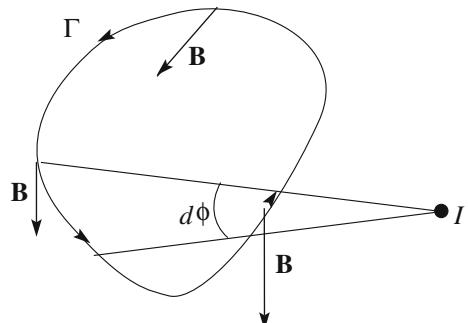
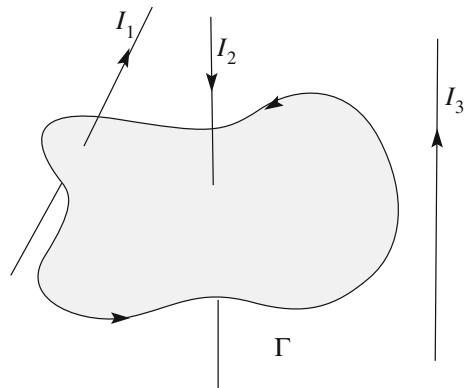


Fig. 6.17 A closed oriented curve and three currents



where I_{link} is the sum of the currents linked to the curve. In the example of Fig. 6.17, it is $I_{\text{link}} = I_1 - I_2$. Note that I_2 has the minus sign because it sees the curve in a clockwise direction and that I_3 does not contribute because it is not linked.

Let us specify the meaning of the “linked” adjective. As a matter of fact, the current carrying wires cannot be infinitely long lines such as those we have drawn. In any case, the circuit must be closed by a return wire. The linked currents are those linked with Γ , like the links of a chain.

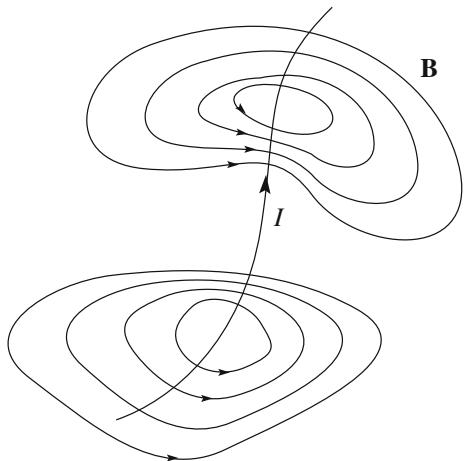
In 1826, André-Marie Ampère (France, 1775–1836) discovered, through a series of fundamental experiments, that Eq. (6.25) has a completely general validity for any shape of the current carrying circuits, as long as the currents are steady. This is called the Ampère circuital law. We assume it to be the fundamental law of magnetostatics.

In conclusion, every system of steady currents is the origin of a magnetic field. The field lines do not have origins (sources) or ends (sinks), but rather they form closed loops, which link with the current lines. Figure 6.18 shows an example for a single current in a wire. More complicated situations arise in the presence of several currents. Note that, in some cases, lines coming from and going to infinite may exist.

If the currents are not carried by wires, and are consequently localized in them, we must describe the situation in terms of the current density \mathbf{j} . Let us generalize Eq. (6.19) for this case. Given the oriented curve Γ , we choose a surface Σ having Γ as its boundary curve. We orient Σ in such a way that the positive direction of its unit normal vector \mathbf{n} sees the positive direction of Γ as being counter-clockwise. I_{link} , the total current through Σ , is then

$$I_{\text{link}} = \int_{\Sigma} \mathbf{j} \cdot \mathbf{n} d\Sigma.$$

Fig. 6.18 A generic stationary current and the magnetic field lines it generates



This equation makes sense only if the integral on the right-hand side does not depend on the choice of Σ . Let us show that to be true, considering two surfaces Σ and Σ' of boundary Γ , as in Fig. 6.19. We want to show that

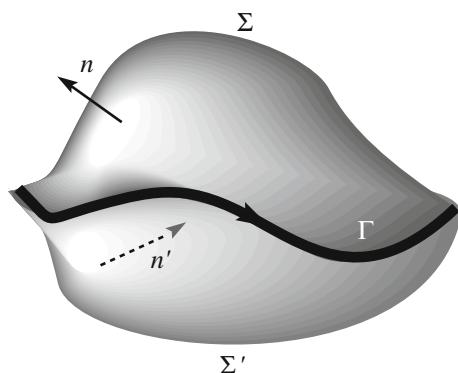
$$\int_{\Sigma} \mathbf{j} \cdot \mathbf{n} d\Sigma = \int_{\Sigma'} \mathbf{j} \cdot \mathbf{n}' d\Sigma'$$

or that

$$\int_{\Sigma} \mathbf{j} \cdot \mathbf{n} d\Sigma - \int_{\Sigma'} \mathbf{j} \cdot \mathbf{n}' d\Sigma' = 0.$$

We notice that Σ and Σ' taken together make a closed surface. If we call \mathbf{n}_{out} the outgoing unit vector of this surface, we see that it coincides with \mathbf{n} on Σ and

Fig. 6.19 Two surfaces with the same boundary



with $-\mathbf{n}'$ on Σ' . Hence, the left-hand side of the last equation is just the flux of \mathbf{j} flowing out from $\Sigma + \Sigma'$. Being that \mathbf{j} is a solenoidal field under steady conditions, its flux flowing out any closed surface is zero. Let us look at this. Indeed, the charge conservation equation $\nabla \cdot \mathbf{j} = -\partial\rho/\partial t$ under stationary conditions, $\rho = \text{const}$, becomes $\nabla \cdot \mathbf{j} = 0$. Calling V the volume enclosed by $\Sigma + \Sigma'$ and integrating on V the divergence $\nabla \cdot \mathbf{j}$, we have $0 = \int_V \nabla \cdot \mathbf{j} dV = \int_{\Sigma + \Sigma'} \mathbf{j} \cdot \mathbf{n}_{\text{out}}$.

We can understand the situation visually, by considering that the lines of current are closed curves. Consequently, the numbers of lines entering into and coming out of a closed surface are equal. The flux is zero. Note, however, that this is true only under stationary conditions, in which \mathbf{j} and ρ do not depend on time. Under dynamical conditions, contrastingly, the Ampère circuital law cannot hold, as we shall see in Chap. 10.

Summarizing, we can write the Ampère circuital law in the form

$$\oint_{\Gamma} \mathbf{B} \cdot d\mathbf{s} = \mu_0 \int_{\Sigma} \mathbf{j} \cdot \mathbf{n}_{\text{out}} d\Sigma. \quad (6.26)$$

where Γ is an oriented closed curve and Σ is any surface having Γ as its boundary oriented with its positive normal vector seeing the positive direction of Γ as counter-clockwise.

The law can be expressed in a differential form as well. To do that, we use the Stokes theorem

$$\oint_{\Gamma} \mathbf{B} \cdot d\mathbf{s} = \int_{\Sigma} (\nabla \times \mathbf{B}) \cdot \mathbf{n} d\Sigma$$

and, considering that Σ is arbitrary, we have

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j}. \quad (6.27)$$

Equations (6.20) and (6.22) in integral form, or the equivalent Eqs. (6.21) and (6.27) in differential form, are the fundamental equations of magnetostatics. They allow us, in principle, to calculate the magnetic field produced by any stationary currents distribution. We anticipate that Eqs. (6.20) and (6.21) are also valid under dynamic conditions, while Eqs. (6.22) and (6.27) are not, needing to be generalized.

6.7 Applications of Ampère's Law

The calculation of the magnetic field generated by a given system of currents generally requires using both Eqs. (6.20) and (6.22). However, in particularly symmetric cases, the problem can be solved using the Ampère circuital law alone.

The approach is similar to that of Sect. 1.13, in which the symmetry of the problem allowed us to find the electric field only using only the Gauss law.

The first example is a toroidal solenoid. A solenoid is a coil wound into a tightly packed helix. The two most interesting geometries are the cylinder and the torus (Fig. 6.20). The axes of the solenoid are, respectively, a straight line and a circle. Let us start by considering a torus with a square cross section.

Let R_1 and R_2 be the radii of the torus ($R_1 < R_2$), N the number of loops and I the current intensity. The loops of a solenoid are densely packed, namely the pitch of the helix is small compared with the length of a loop. The symmetry of the problem and the fact that the lines of \mathbf{B} are closed suggest that they should be circles concentric with and, in planes, parallel to the torus. Symmetry also requires the magnitude B to depend only on the distance r from the center.

It is easy to see that the field is zero outside the solenoid.

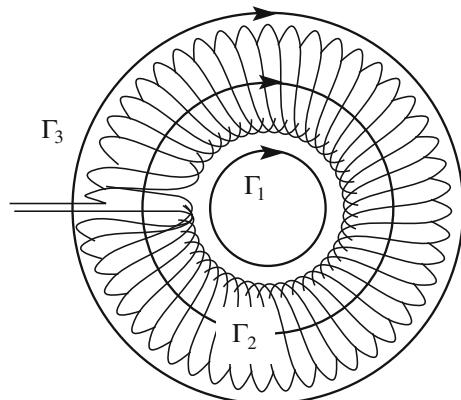
To see that, we apply the Ampère law to a circle as Γ_1 in Fig. 6.20, of radius r_1 smaller than R_1 . Being that $I_{\text{link}} = 0$, we have $\oint_{\Gamma_1} \mathbf{B} \cdot d\mathbf{s} = 2\pi r_1 B = 0$, and consequently $B = 0$. The linked current $I_{\text{link}} = 0$, also for Γ_3 in Fig. 6.20, of radius $r_3 > R_3$ because the enclosed circle is crossed by N currents in one direction and N equal currents in the opposite one. Hence, $B = 0$ here as well.

To obtain the field $B(r)$ inside the solenoid, we apply the Ampère law to the circle Γ_2 of radius r , with $R_1 < r < R_2$. We obtain $\oint_{\Gamma_2} \mathbf{B} \cdot d\mathbf{s} = 2\pi r B = \mu_0 N I$. Hence, we have

$$B(r) = \mu_0 \frac{NI}{2\pi r}. \quad (6.28)$$

We see that the field intensity decreases for r increasing from R_1 to R_2 as $1/r$. In addition, the strength is higher for larger numbers of coils and larger current intensities. We also observe that if $R_2 - R_1 \ll R_1$, then $2\pi r$ is the length of the solenoid. If n is the number of loops per unit length, namely $n = N/(2\pi r)$, we have

Fig. 6.20 A toroidal solenoid



$$B(r) \cong \mu_0 n I. \quad (6.29)$$

The loops being very tight, we can think of the torus as a continuous current carrying surface with surface current density that we call k . This is the current intensity per unit length of the solenoid, namely $k = nI$. We can then also write

$$B(r) \cong \mu_0 k. \quad (6.30)$$

We note that, crossing the surface current, moving from inside to outside the solenoid, the tangential component of the field (the unique non-zero, in this case) has a discontinuity equal to $\mu_0 k$. As we shall soon see, this discontinuity exists, in general, when crossing a current sheet.

Let us consider now a rectilinear cylindrical solenoid. We shall assume it to be infinitely long, which is obviously an unrealistic case. We shall not make any assumption as to the shape of the section. Hence, our result will be independent of that. Let n be the number of loops per unit length along the axis (the total number of loops is infinite). We shall again assume the loops to be very dense and the current lines to be circles normal to the axis. We can think about opening the toroidal solenoid we have just considered and extending it to infinite length. We can then conclude that the field should be zero outside the rectilinear solenoid as well. The lines of the field should be straight lines inside the solenoid, parallel to the axis and extending to infinite.

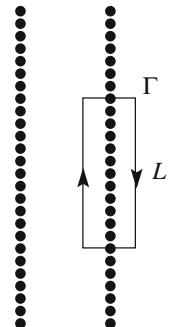
Let us apply the Ampère law to the curve Γ , as shown in Fig. 6.21. The curve is made of two segments of length L parallel to the axis, one inside and one outside the solenoid and two segments normal to the axis. The unique contribution to the circulation integral of \mathbf{B} is on the internal segment L . nL being the number of loops crossing the curve, the Ampère law gives us $BL = \mu_0 nIL$. The field is then

$$B(r) = \mu_0 n I = \mu_0 k \quad (6.31)$$

where $k = nI$ is the surface current density.

We observe that the field inside the solenoid does not depend on the distance from the axis, namely it is uniform.

Fig. 6.21 A linear solenoid and a curve on which to apply the Ampère law



Also, in this case, the discontinuity of the tangential component of \mathbf{B} crossing the current sheet of current density k is $\mu_0 k$.

In practice, the cylindrical solenoids have finite lengths. Equation (6.31) is a reasonable approximation for the field inside a long solenoid, having a length much larger than the base diameter, at points not too close to the extremes. The field outside it is weak but not zero. The field lines are shown in Fig. 6.22.

Let us now consider the unrealistic but logically relevant case of an infinite current-carrying plane, with uniform surface current density \mathbf{k} . Let us choose a reference frame with the z -axis in the direction of \mathbf{k} , the x -axis on the current plane as well, and the y -axis perpendicular to the plane. In this frame, the surface current density has the components $\mathbf{k} = (0, 0, k)$. Considering that the lines of \mathbf{B} should run about the current in a counter-clockwise direction and the symmetry of the problem, we understand that the lines of \mathbf{B} are straight lines in the x direction. More precisely, \mathbf{B} is in the positive x direction on the side of negative y , and in the opposite direction on the side of positive y .

Symmetry requires that the magnitude of \mathbf{B} depend only on the distance from the plane, if any exists. We apply the Ampère law to the line Γ shown in Fig. 6.23. The curve is a rectangle with two sides parallel to x at distances $-y$ and $+y$ from the plane. The Ampère law gives $\oint_{\Gamma} \mathbf{B} \cdot d\mathbf{s} = 2BL = \mu_0 kL$. The magnitude of the field is then

$$B = \frac{\mu_0}{2} k \quad (6.32)$$

Fig. 6.22 Field lines of a cylindrical solenoid

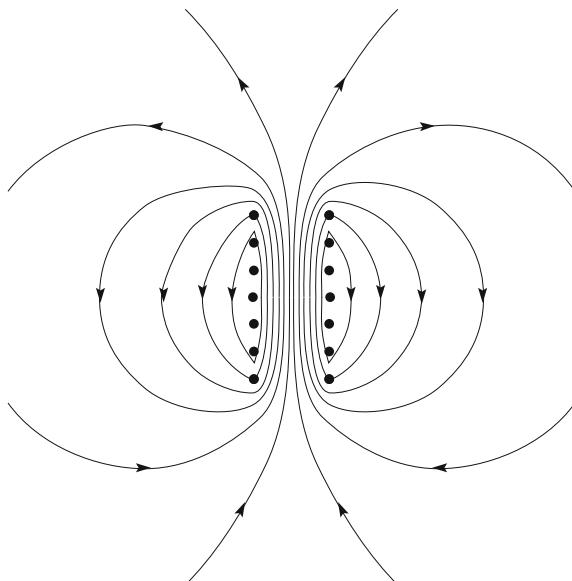
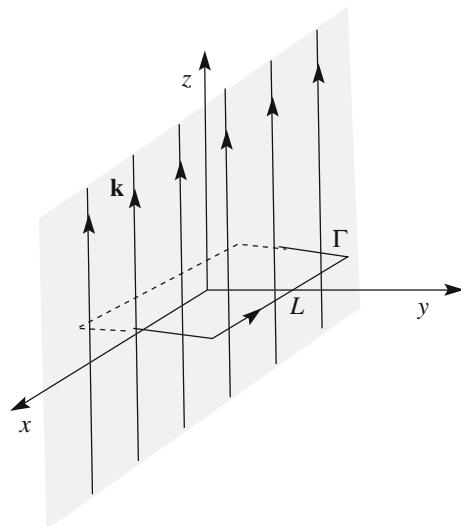


Fig. 6.23 A uniform current-carrying plane surface



The Cartesian components of the field are

$$\mathbf{B} = \left(-\frac{\mu_0}{2}k, 0, 0 \right) \text{ for } y > 0; \quad \left(+\frac{\mu_0}{2}k, 0, 0 \right) \text{ for } y < 0 \quad (6.33)$$

We observe that the field does not depend on the distance from the current plane. We already encountered a similar circumstance for the electric field generated by an infinite charged plane. The reasons are similar. Now, the substantial contributions to the magnetic field at the distance y come from the current density in a circle of radius on the order of y . When the distance increases, the contribution of each unit surface decreases (inversely to the distance), but the number of contributing surface units increases (proportionally to the distance). The two effects balance one another out.

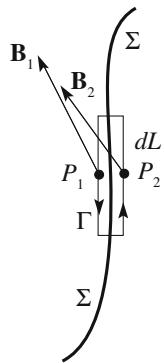
We finally observe that, once more, the tangential component of \mathbf{B} has a discontinuity in crossing the current sheet, of magnitude $\mu_0 k$.

Let us now see that this property is general, namely let us consider the discontinuities of parallel and perpendicular components of the magnetic field crossing a current-carrying surface.

Let Σ be an arbitrary surface carrying a current of surface density \mathbf{k} . We consider the vector \mathbf{k} as being given as a function of the point on the surface. Note that \mathbf{k} is, at every point, tangent to Σ at that point.

Figure 6.24 represents the current surface Σ cut by the plane of the drawing. The current exits normally from the page. Let P_1 and P_2 be two points infinitely close to Σ on the two faces. We analyze the behavior of the tangent component by applying the Ampère law to the curve Γ in Fig. 6.24. The curve is made of two segments of infinitesimal length dL parallel to the surface through P_1 and P_2 , and two perpendicular to the surface. The latter are taken as infinitesimal of an order superior to that of dL in order to give a negligible contribution to the circulation integral.

Fig. 6.24 A current carrying surface and a closed line to which the Ampère law is applied



Let \mathbf{B}_T be the projection of the field on the current plane. The Ampère law gives us $\oint_{\Gamma} \mathbf{B} \cdot d\mathbf{s} = [B_T(P_2) - B_T(P_1)]dL = \mu_0 k dL$. The discontinuity of the tangential component is then

$$\Delta B_T = \mu_0 k. \quad (6.34)$$

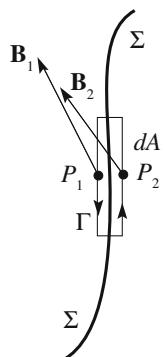
Let us now consider the behavior of the component of the field \mathbf{B} normal to the surface Σ , which we call B_N . Figure 6.25 again represents a section of Σ , but the small rectangle through P_1 and P_2 is now the section of a closed surface to which we apply the Gauss law. The argument is the same one we made for electrostatics.

The closed surface is a box with two infinitesimal faces parallel to Σ of surface dA and a lateral surface normal to Σ taken to be infinitesimal of an order superior to that of dA , in order to give a negligible contribution to the outgoing flux. The outgoing flux of \mathbf{B} is zero, namely $B_N(P_2)dA - B_N(P_1)dA = 0$. Then, we have

$$\Delta B_N = 0. \quad (6.35)$$

The component of the \mathbf{B} field normal to the surface is continuous.

Fig. 6.25 A current-carrying surface and a box to which the Gauss law is applied



6.8 The Vector Potential

As we have seen, under stationary conditions, the magnetic field is originated by electric currents according the two equations

$$\nabla \cdot \mathbf{B} = 0 \quad (6.36)$$

and

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j}. \quad (6.37)$$

This is somewhat similar to electrostatics, in which the electric field generated by stationary charges obeys the two equations

$$\nabla \cdot \mathbf{E} = \rho / \epsilon_0 \quad (6.38)$$

and

$$\nabla \times \mathbf{E} = 0. \quad (6.39)$$

In electrostatics, being that the curl of the field is zero, we can introduce a scalar field ϕ , namely the potential, such as $\mathbf{E} = -\nabla\phi$. We cannot define a similar potential for the magnetic field, because its curl is not zero. However, the field has zero divergence. Considering that the divergence of the curl of any vector field is zero, we can introduce a vector field \mathbf{A} , the curl of which is \mathbf{B} , namely

$$\mathbf{B} = \nabla \times \mathbf{A}. \quad (6.40)$$

\mathbf{A} is called the *vector potential*. The physical dimensions of \mathbf{A} are those of a magnetic field times a length (because the gradient has the dimension of an inverse length). Consequently, the vector potential is measured in tesla times meters, T m. In addition, considering that the dimensions of \mathbf{B} are a force divided by a velocity and a charge, we can write the dimensional equation as

$$[\mathbf{A}] = [\mathbf{BL}] = [FTL^{-1}Q^{-1}L] = [FTQ^{-1}] = [pQ^{-1}],$$

where, on the right-hand side, we have recalled that the linear momentum, p , has the dimensions of a force times a time. In conclusion, the vector potential is dimensionally a momentum per unit charge. From this point of view, \mathbf{A} is analogous to the scalar potential, which is energy per unit charge. In Sect. 10.11, we shall see that the analogy is not only formal but has a deep physical meaning.

All the effects measured in a field are due to forces. In electrostatics, the force is due to the electric field. A consequence is that the scalar potential is defined only up to an additive constant. Similarly, the magnetic forces that we measure are due to \mathbf{B} . Hence, if instead of \mathbf{A} , we take an \mathbf{A}' that is such so as to give the same \mathbf{B} , \mathbf{A}' is also

a valid vector potential for the given \mathbf{B} . This happens if $\nabla \times \mathbf{A}' = \nabla \times \mathbf{A}$, or $\nabla \times (\mathbf{A}' - \mathbf{A}) = 0$. Namely, the curl of the vector $\mathbf{A}' - \mathbf{A}$ is identically zero. As such, we can express it as the gradient of an arbitrary scalar function, namely

$$\mathbf{A}' - \mathbf{A} = \nabla \psi. \quad (6.41)$$

In conclusion, the vector potential \mathbf{A} is defined up to the gradient of an arbitrary scalar function. This quite large arbitrariness of choice corresponds to the fact that Eq. (6.40) only fixes the curl, while the sources of \mathbf{A} , namely its divergence, are arbitrary. We can then choose any value for $\nabla \cdot \mathbf{A}$ without changing \mathbf{B} . We shall exploit this arbitrariness to make the equations as simple as possible. This operation is called ‘choosing a gauge’ and also ‘gauge fixing’. The convenient choice in magnetostatics is called the Coulomb gauge, which is taking the divergence of the vector potential equal to zero, namely

$$\nabla \cdot \mathbf{A} = 0 \quad (6.42)$$

Note that choosing the gauge does not yet completely determine the vector potential. Some arbitrariness remains. We can, for example, add a uniform vector field to \mathbf{A} .

As with \mathbf{B} , the vector potential \mathbf{A} is generated by the currents. To see its relation with the current density, we just substitute Eq. (6.40) in Eq. (6.37), obtaining $\nabla \times (\nabla \times \mathbf{A}) = \mu_0 \mathbf{j}$. Recalling the vector identity $\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$ and using Eq. (6.42) (we made that choice on purpose), we get

$$\nabla^2 \mathbf{A} = -\mu_0 \mathbf{j} \quad (6.43)$$

These are three equations, one for each component of the vectors, namely $\nabla^2 A_x = -\mu_0 j_x$ and two similar ones for the x and y components. Contents apart, they are the same equation that holds for the electrostatic potential, namely

$$\nabla^2 \phi = -\frac{\rho}{\epsilon_0}.$$

We know its solution, namely

$$\phi(\mathbf{r}_1) = \frac{1}{4\pi\epsilon_0} \int \frac{\rho(\mathbf{r}_2)}{r_{12}} dV_2.$$

As equal equations have equal solutions, we also know the solution to Eq. (6.43). We just have to change ρ/ϵ_0 in $\mu_0 j_x$ etc., obtaining

$$A_x(\mathbf{r}_1) = \frac{\mu_0}{4\pi} \int \frac{j_x(\mathbf{r}_2)}{r_{12}} dV_2.$$

or, in vector form

$$\mathbf{A}(\mathbf{r}_1) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{j}(\mathbf{r}_2)}{r_{12}} dV_2 \quad (6.44)$$

An important case is one in which the current is in a wire circuit. Let Σ be a normal section of the circuit and I the current intensity. Let Γ be the geometrical curve of the circuit, oriented as the current. As shown in Fig. 6.26, let O be the origin of the axes and $d\mathbf{s}$ a generic element of Γ , at the position vector \mathbf{r}_2 . Let P be the point at \mathbf{r}_1 where we want to express the vector potential.

The current density \mathbf{j} is different from zero only inside the wire. We take an infinitesimal volume dV_2 inside the wire of length $d\mathbf{s}$ in the wire direction and base $d\Sigma$ normal to the wire. We perform the integral in Eq. (6.44), first integrating on $d\Sigma$ along the entire section Σ and then on $d\mathbf{s}$ along the complete circuit. We now suppose the wire to be thin and P not very close to the wire, in order that $r_{12} = |\mathbf{r}_2 - \mathbf{r}_1|$ is much larger than the diameter of the wire. In this hypothesis, r_{12} is practically a constant when we integrate on Σ . In addition, we take into account that \mathbf{j} and $d\mathbf{s}$ have the same direction. We can then write

$$\mathbf{A}(\mathbf{r}_1) = \frac{\mu_0}{4\pi} \oint_{\Gamma} ds \int_{\Sigma} \frac{\mathbf{j}(\mathbf{r}_2)}{r_{12}} d\Sigma = \frac{\mu_0}{4\pi} \oint_{\Gamma} \frac{d\mathbf{s}}{r_{12}} \int_{\Sigma} j(\mathbf{r}_2) d\Sigma.$$

The integral on Σ on the right-hand side is just the current intensity I , and finally, we have

$$\mathbf{A}(\mathbf{r}_1) = \frac{\mu_0 I}{4\pi} \oint_{\Gamma} \frac{d\mathbf{s}}{r_{12}} \quad (6.45)$$

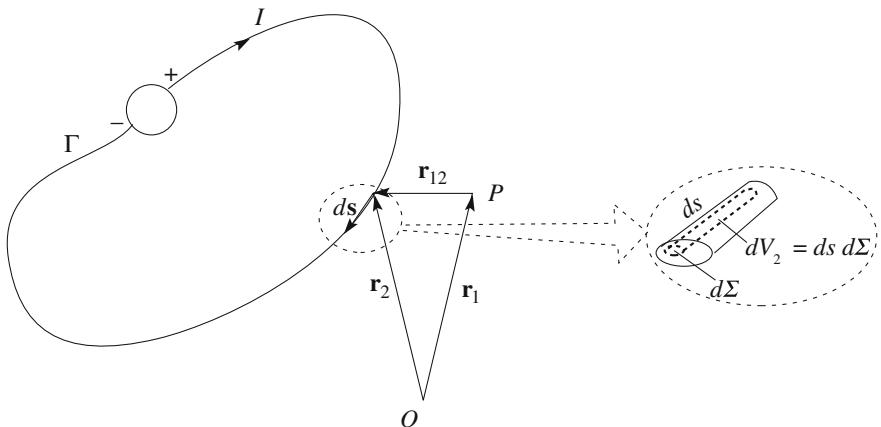


Fig. 6.26 Calculating the vector potential for a current-carrying wire

In conclusion, to get the vector potential of a current-carrying wire, at distances that are large compared with the diameter of the wire, we have to calculate three line integrals, one for each component, of Eq. (6.45). Clearly, once we have \mathbf{A} , we obtain the magnetic field by taking its curl.

Equation (6.40), which is the definition of the vector potential, and the choice of the gauge (6.42) are local relations. They can be expressed in integral form as well. The integral expressions are useful in certain circumstances.

Let Γ be an arbitrary closed oriented curve and Σ a surface bound by Γ and coherently oriented. Let us calculate the flux of \mathbf{B} through Σ and apply the Stokes theorem:

$$\int_{\Sigma} \mathbf{B} \cdot \mathbf{n} d\Sigma = \int_{\Sigma} \nabla \times \mathbf{A} \cdot \mathbf{n} d\Sigma = \oint_{\Gamma} \mathbf{A} \cdot d\mathbf{s},$$

or

$$\oint_{\Gamma} \mathbf{A} \cdot d\mathbf{s} = \int_{\Sigma} \mathbf{B} \cdot \mathbf{n} d\Sigma = \Phi_B. \quad (6.46)$$

Namely, the circulation of \mathbf{A} around any closed curve is equal to the flux of \mathbf{B} linked to it.

The gauge condition Eq. (6.42) can similarly be cast in integral form, by calculating the flux of \mathbf{A} coming out of an arbitrary closed surface, which we call Σ , and applying the Gauss theorem. We get

$$\oint_{\Sigma} \mathbf{A} \cdot \mathbf{n} d\Sigma = 0. \quad (6.47)$$

Namely, the flux of the vector potential coming out of any closed surface is zero.

In conclusion, the vector field \mathbf{A} is solenoidal; its lines are closed curves, or curves coming from and going to infinite. There are no sources of \mathbf{A} , namely points from which the curves of \mathbf{A} exit or into which they enter. The lines of \mathbf{A} are analogous to the lines of \mathbf{B} , with Φ_B in place of the currents; they tend to embrace the lines of \mathbf{B} .

QUESTION Q 6.4. With the Coulomb gauge, can closed lines exist for the vector potential? Can lines coming out of a point? \square

6.9 The Vector Potential in Simple Cases

Equation (6.45) allows us to calculate the vector potential \mathbf{A} generated by any given system of currents. Once we have \mathbf{A} , we obtain the magnetic field \mathbf{B} by finding its curl. The difficulty of the actual calculations depends on the shape of the circuit.

We shall consider a few simple cases, taking advantage, as much as possible, from the analogy with electrostatic charge distributions $\rho(\mathbf{r}_2)$, of which we already know the $\phi(\mathbf{r}_1)$. As a matter of fact, each of these cases provides us with a ready solution to a magnetostatic problem. We obtain the expression of a component of the vector potential, say $A_x(\mathbf{r}_1)$, for example, generated by a current density of x -component $j_x(\mathbf{r}_2)$ from the known expression of $\phi(\mathbf{r}_1)$ of the analogous $\rho(\mathbf{r}_2)$, with the substitution

$$\rho/\varepsilon_0 \rightarrow \mu_0 j_x. \quad (6.48)$$

Consider, as a first example, a current-carrying straight wire. Let I be the current intensity. We choose the reference system shown in Fig. 6.27, with the z -axis laying on the current in its positive direction. The current density \mathbf{j} has only one non-zero component, namely j_z . Consequently, the only non-zero component of the vector potential is A_z .

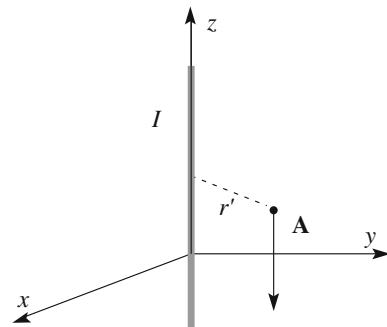
The electrostatic analogy is a uniformly-charged straight wire. Let λ be its linear charge density. Its potential at the distance r' from the z -axis is $\phi(r') = -\lambda/(2\pi\varepsilon_0) \ln r'$. The charge ΔQ in a segment on length Δl of the wire is, in terms of the volumetric density ρ , equal to $\rho S \Delta l$, and, in terms of the linear density, equal to $\lambda \Delta l$. Then, it is $\rho = \lambda/S$. We see that the substitution of Eq. (6.48) becomes for linear charge distributions $\lambda/\varepsilon_0 \rightarrow \mu_0 I$. We obtain

$$A_z(r') = -\frac{\mu_0 I}{2\pi} \ln r', \quad A_x(r') = 0, \quad A_y(r') = 0. \quad (6.49)$$

The vector potential lines are straight lines parallel to the current. The magnitude of \mathbf{A} grows with the distance from the current, slowly but without limit. The reason for this strange behavior is exactly the same as that for the analogous electrostatic potential. The unrealistic situation is a consequence of considering a current that extends to infinity.

We can now obtain the known result for \mathbf{B} by calculating the $\nabla \times \mathbf{A}$. We immediately see that $B_z = 0$. For the other components, we have

Fig. 6.27 A current-carrying straight wire and its vector potential



$$B_x = \frac{\partial A_z}{\partial y} - 0 = -\frac{\mu_0 I}{2\pi} \frac{\partial \ln r'}{\partial y} = -\frac{\mu_0 I}{2\pi} \frac{y}{r'^2}$$

$$B_y = 0 - \frac{\partial A_z}{\partial x} = \frac{\mu_0 I}{2\pi} \frac{\partial \ln r'}{\partial x} = \frac{\mu_0 I}{2\pi} \frac{x}{r'^2}$$

We see that B_x is proportional to $-y$ and B_y is proportional to x with the same proportionality constant. Hence, $\mathbf{B} \cdot \mathbf{r}' = 0$, namely the direction of \mathbf{B} is perpendicular to \mathbf{r}' at any point. We earlier found that the lines of \mathbf{B} are circles normal to the current. Its absolute value decreases at an increasing distance from the wire as $B(r') = \frac{\mu_0 I}{2\pi} \frac{1}{r'}$.

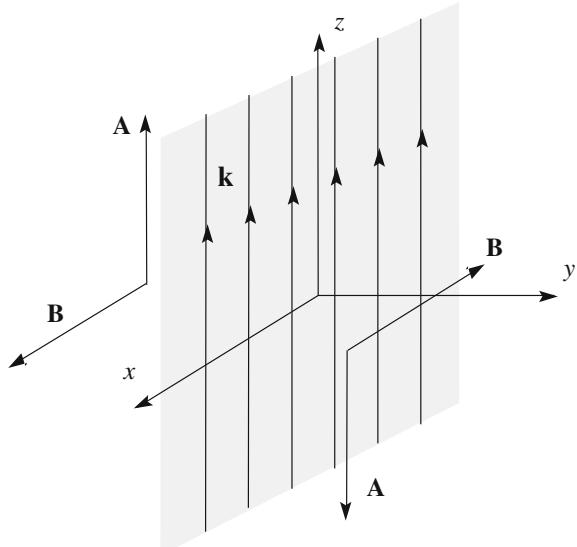
Let us now consider a current-carrying plane with constant and uniform surface density k . We choose the reference frame, as in Fig. 6.28, with the z -axis in the plane in the positive direction of the current and the x -axis in the plane as well. Being that k_x and k_y are identically zero, the only non-zero component of \mathbf{A} is A_z .

The electrostatic analogy is a uniformly charged plane with surface density σ . We know that its potential depends only on the absolute distance $|y|$ from the plane and that the potential is $\phi(y) = -\sigma|y|/(2\epsilon_0)$. To use the substitution of Eq. (6.48), we need a volume charge density ρ . We must think of the surface as having a small thickness s . Then, the charge ΔQ in an area ΔS of the surface is $\rho s \Delta S$ when thought of as a volume density and $\sigma \Delta S$ when thought of as a surface density. Similarly, the current intensity ΔI crossing a section $s \Delta x$ is $j_z s \Delta x$ if thought of as a volume current and $k \Delta x$ if thought of as a surface current. Hence, we have $j_z s = k$.

In conclusion, the substitution of Eq. (6.48) becomes, for two-dimensional charge distributions,

$$\sigma/\epsilon_0 \rightarrow \mu_0 k.$$

Fig. 6.28 A current-carrying plane and its vector potential



The vector potential components are then

$$A_x(x, y, z) = 0, \quad A_y(x, y, z) = 0, \quad A_z(x, y, z) = -\frac{\mu_0 k}{2} |y|. \quad (6.49)$$

Also, in this case, the vector potential has the same direction as the current. As for the magnetic field, we immediately see that $B_y = B_z = 0$ and that $B_x = -\mu_0 k/2$ for $y > 0$, $B_x = +\mu_0 k/2$ for $y < 0$, which is the result we know.

Consider now an infinite cylindrical solenoid. We consider it to be a cylindrical surface of radius R carrying a surface current density \mathbf{k} on its lateral surface. The lines of current are circles normal to the axis. In practice, the solenoid is made of loops of wire, n per unit length in number, and carrying the current intensity I , such that $nI = k$.

We choose the reference frame of Fig. 6.29 with the z -axis on the axis of the cylinder positively oriented to see the current circulating counter-clockwise.

We know that the magnetic field is zero outside the cylinder. The lines of the vector potential \mathbf{A} revolve around the cylinder in a manner similar to the lines of the field \mathbf{B} produced by a cylindrical current. Symmetry suggests to us that the lines of \mathbf{A} are circumferences normal to the cylinder with centers on the axis and that the magnitude of \mathbf{A} depends only on the distance, which we call r' , from the axis.

Consider one of these circles, such as Γ_1 in Fig. 6.29b, external to the cylinder, namely having radius $r' > R$, and apply to it Eq. (6.46). The circulation integral of \mathbf{A} is $2\pi r' A(r')$, because \mathbf{A} is everywhere tangent to the curve and with the same magnitude. We then have $2\pi r' A(r') = \Phi_B = \pi R^2 \mu_0 k$, where we have considered that B is uniform in the cylinder with magnitude $B = \mu_0 k$. We obtain

$$A(r') = \frac{\mu_0 R^2 k}{2r'} \quad \text{for } r' > R. \quad (6.50)$$

It is a simple exercise to show that, outside the solenoid, both the divergence and the curl of the vector \mathbf{A} are zero.

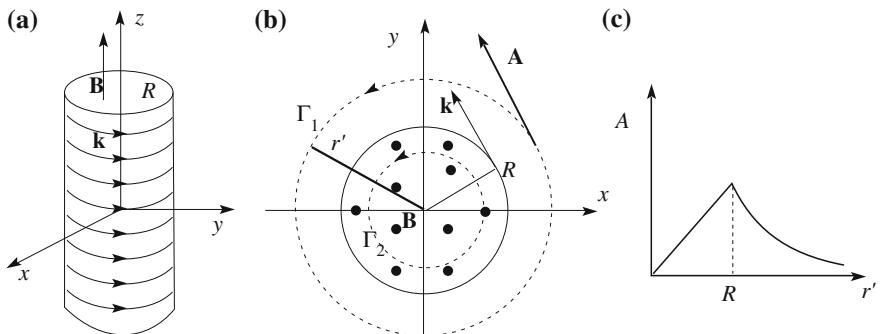


Fig. 6.29 A cylindrical solenoid and its vector potential

To find \mathbf{A} inside the solenoid, we now apply Eq. (6.46) to the circle Γ_2 in Fig. 6.29b internal to the cylinder, namely having radius $r' < R$. We obtain $2\pi r' A(r') = \Phi_B = \pi r'^2 \mu_0 k$, which gives us

$$A(r') = \frac{\mu_0 k}{2} r' \quad \text{for } r' < R. \quad (6.51)$$

We see that A is zero on the axis, increases linearly with the distance from the axis inside the solenoid and decreases as the inverse of that distance outside, as shown in Fig. 6.29c.

6.10 The Ampère-Laplace Law

We have discussed how to calculate the vector potential \mathbf{A} for a given system of currents and then how to obtain \mathbf{B} by taking the curl of \mathbf{A} . We shall now see how to express \mathbf{B} directly in terms of the currents that generate it.

Let us consider a generic circuit carrying the steady current I , as in Fig. 6.30.

The magnetic field \mathbf{B} at the generic point $P_1 \equiv (x_1, y_1, z_1)$ is obtained by integration on the volume of the circuit:

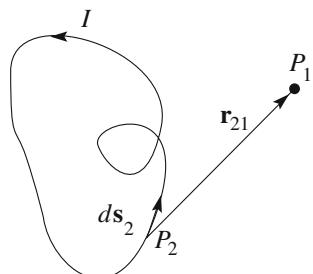
$$\mathbf{B}(P_1) = \nabla_1 \times \mathbf{A}(P_1) = \nabla_1 \times \int \frac{\mathbf{j}(P_2) dV_2}{r_{21}}$$

where $P_2 \equiv (x_2, y_2, z_2)$ is the point running on the integration domain and $dV_2 = dx_2 dy_2 dz_2$. The footer in ∇_1 is to indicate that the partial derivatives are in x_1, y_1, z_1 rather than in x_2, y_2, z_2 .

Let us start with the x -component of \mathbf{B} . We take the operator ∇_{1x} inside the integral and remember that the components of \mathbf{j} do not depend on P_1 . We have

$$B_x(P_1) = \frac{\partial A_z}{\partial y_1} - \frac{\partial A_y}{\partial z_1} = \frac{\mu_0}{4\pi} \int \left[j_z(P_2) \frac{\partial(1/r_{21})}{\partial y_1} - j_y(P_2) \frac{\partial(1/r_{21})}{\partial z_1} \right] dV_2.$$

Fig. 6.30 A generic steady current circuit



Recalling that $1/r_{21} = \left[(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2\right]^{-1/2}$ and differentiating, we get

$$B_x(P_1) = \frac{\mu_0}{4\pi} \int \left[-j_z(P_2) \frac{y_1 - y_2}{r_{21}^3} + j_y(P_2) \frac{z_1 - z_2}{r_{21}^3} \right] dV_2 = \frac{\mu_0}{4\pi} \int \left[\frac{\mathbf{j}(P_2) \times \mathbf{r}_{21}}{r_{21}^3} \right]_x dV_2$$

and similar expressions for the other two components. Finally, calling \mathbf{u}_{21} the unit vector of \mathbf{r}_{21} , we can write

$$\mathbf{B}(P_1) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{j}(P_2) \times \mathbf{r}_{21}}{r_{21}^3} dV_2 = \frac{\mu_0}{4\pi} \int \frac{\mathbf{j}(P_2) \times \mathbf{u}_{21}}{r_{21}^2} dV_2. \quad (6.52)$$

If the currents are confined in wire circuits and we are interested in the field that is not very close to any wire, we can operate as we did in Sect. 6.7. In the case of a single circuit Γ carrying the current I , we integrate first on the section of the wire and pose $\int j d\Sigma = I$. We note that the integration element $d\mathbf{s}_2$ has the same direction as $\mathbf{j}(P_2)$ and write

$$\mathbf{B}(P_1) = \frac{\mu_0}{4\pi} I \int \frac{d\mathbf{s}_2 \times \mathbf{u}_{21}}{r_{21}^2} = -\frac{\mu_0}{4\pi} I \int \frac{\mathbf{u}_{21} \times d\mathbf{s}_2}{r_{21}^2}. \quad (6.53)$$

This expression is called the Ampère-Laplace law. The law can also be stated by saying that each element $d\mathbf{s}_2$ of the circuit produces an elementary magnetic field

$$d\mathbf{B}(P_1) = -\frac{\mu_0}{4\pi} I \frac{\mathbf{u}_{21} \times d\mathbf{s}_2}{r_{21}^2}. \quad (6.54)$$

This equation is called, by some authors, the 1st Laplace law.

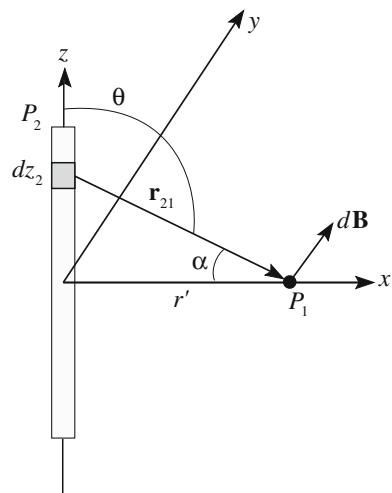
The extension of Eq. (6.53) to more than one circuit is obvious. One only needs to add the integrals relative to each circuit.

6.11 Examples of Magnetic Field Calculations

In this section, we give a few examples of magnetic field calculations, using the Ampère-Laplace law. We remind the reader that the results will be valid at distances from the current-carrying wire much larger than its diameter. The first example is, once more, a straight current.

We choose a reference frame with the z -axis in the wire oriented as the current and a x -axis normal to it, as in Fig. 6.31. Symmetry requires that the field intensity should depend only on the distance from the wire, which we shall call r' . We can then calculate the field at a point of the x -axis, as P_1 in the figure, without losing generality.

Fig. 6.31 Calculating the magnetic field of a straight steady current



Let $\mathbf{u}_{21} = (\sin \theta, 0, \cos \theta)$ be the unit vector and $ds_2 = (0, 0, dz_2)$ the line element of the wire. Equation (6.54) gives, for the magnitude of the field,

$$dB = \frac{\mu_0 I dz_2 \sin \theta}{4\pi r_{21}^2} = \frac{\mu_0 I dz_2 \cos \alpha}{4\pi r_{21}^2}.$$

We now express the geometrical quantities r_{21} and dz_2 as functions of α , on which we shall integrate. We have $r_{21} = r'/\cos \alpha$ and $dz_2 = (r'/\cos^2 \alpha) d\alpha$, and we write

$$dB = \frac{\mu_0 I \cos \alpha}{4\pi r'} d\alpha.$$

Integrating on the wire from $-\infty$ to $+\infty$ means integrating on α from $-\pi/2$ to $+\pi/2$. We have

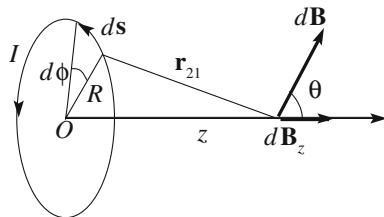
$$B = \frac{\mu_0 I}{4\pi r'} \int_{-\pi/2}^{+\pi/2} \cos \alpha d\alpha = \frac{\mu_0 I}{4\pi r'},$$

which is the result we know.

A circular current loop is a symmetric and geometrically simple configuration. Contrary to intuition, the calculation of its magnetic field at a generic point is very difficult, involving the so-called elliptic integrals. However, things are easy on the axis of the loop.

We choose a reference frame with the origin in the center of the loop, the z -axis perpendicular to the loop oriented to see the direction of the current as counter-clockwise, as in Fig. 6.32. Let R be the radius of the loop. We calculate the field at a generic point P_1 of the z -axis, where symmetry requires the field to have

Fig. 6.32 Calculating the magnetic field on the axis of a steady circular current loop



the direction of the axis. We can then consider only the contribution to the z -component of the field for each loop element $d\mathbf{s}$, namely

$$dB = \frac{\mu_0 I ds}{4\pi r_{21}^2} \cos \theta = \frac{\mu_0 I R d\phi}{4\pi R^2 + z^2} \frac{R}{\sqrt{R^2 + z^2}} = \frac{\mu_0 I}{4\pi} \frac{R^2}{(R^2 + z^2)^{3/2}} d\phi.$$

By integration, we now obtain the z -component of the field, which is the only non-zero one

$$B_z = \frac{\mu_0 I}{2} \frac{R^2}{(R^2 + z^2)^{3/2}}. \quad (6.55)$$

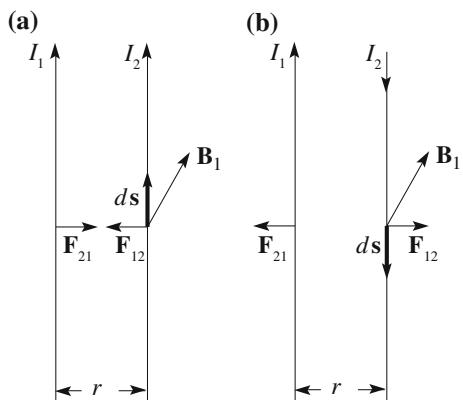
6.12 Force Between Two Straight Steady Currents and the Ampere

If we lay down two straight wires carrying the steady currents of intensities I_1 and I_2 parallel to one another, we observe that they attract each other if the currents have the same direction, and repel if the directions are opposed. Figure 6.32 shows the situation. This fact was first observed by Ampère in 1825 during one of his fundamental experiments.

The forces acting on the wires are equal and opposite, proportional directly to the product of the current intensities and inversely to the distance, which we call r , between the wires. We explain these observations as follows. The current in the first wire generates a magnetic field, whose lines are circles normal to and centered on the wire. The magnitude of the field is $B_1 = \frac{\mu_0 I_1}{4\pi r}$. The positive direction of \mathbf{B}_1 is such that an observer laying on the current with the positive direction entering through his feet and flowing out of his head sees the field lines circulate in a counter-clockwise direction, as in Fig. 6.33.

As a consequence, the generic segment $d\mathbf{s}$ of the second wire is acted upon by a force $d\mathbf{F} = I_2 d\mathbf{s} \times \mathbf{B}_1$, as given by the 2nd Laplace law [Eq. (6.4)]. As \mathbf{B}_1 and $d\mathbf{s}$ are perpendicular, we have $dF = I_2 ds B_1$. The force *per unit length* on wire 2 is then

Fig. 6.33 Forces between two parallel straight steady currents



$$\frac{F_{12}}{l} = I_2 B_1 = \frac{\mu_0 I_1 I_2}{2\pi r}. \quad (6.56)$$

As we said, the force is attractive in the case of parallel currents, as shown in Fig. 6.33a, and repulsive in the case of antiparallel currents, as shown in Fig. 6.33b.

Equation (6.56) is also important, because it is used for the definition of the electromagnetic base unit in the SI, which is the ampere. The ampere is defined by stating by definition the value of the constant $\mu_0/2\pi$, namely

$$\frac{\mu_0}{2\pi} = 2 \times 10^{-7}. \quad (6.57)$$

The official definition of the ampere in the SI is:

The ampere is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed 1 m apart in vacuum, would produce between these conductors a force equal to 2×10^{-7} N per meter of length.

As we already mentioned, the unit of electric charge, the coulomb, is a derived unit, it is the charge going through a section of a wire carrying a current of one ampere (1 A) in one second (1 s).

QUESTION Q 6.5. Two straight parallel wires 0.5 m long placed at 10 cm distance carry a current of 1 A in opposite directions. What is the magnetic force between them and how is it directed? □

6.13 The Magnetic Dipole

The term ‘magnetic dipole’ originated historically, when it was believed that magnetic charges existed, to mean a pair equal and opposite of those charges, or poles, at a certain distance, in analogy with the electric dipole. We know that

magnetic charges do not exist, but we still use the name to mean a planar loop carrying a steady current of small dimensions. “Small” means that we are interested in the magnetic field it produces at distances large compared to the geometric dimensions of the loop. In addition, when we consider the dipole in an external magnetic field, we shall assume that field to vary only a little, if any, on the dimensions of the dipole. The name “dipole” is due to the fact that, as we shall now see, its magnetic field \mathbf{B} is similar to the electric field \mathbf{E} of an electric dipole. More generally, the behavior of a magnetic dipole is similar to that of an electric dipole.

Let us consider a rectangular loop of sides a and b carrying the steady current I . Let us choose the origin of the reference frame at its center, the z -axis normal to the plane of the loop and oriented to see the positive direction of the current as counter-clockwise, and the x and y axes parallel to the sides of the loop, as in Fig. 6.34a. We start by calculating the vector potential \mathbf{A} , at a point P of coordinates (x, y, z) at a distance from the loop that is large compared to a and b .

We begin with the A_x component, corresponding to the j_x component of the current density. We start from the scalar potential of a charge distribution geometrically similar to j_x , as in Fig. 6.34b and apply the substitution $\rho(x, y, z)/\epsilon_0 \rightarrow \mu_0 j_x(x, y, z)$. The current intensity I is the current density j_x times the section of the wire. Hence, the equivalent charge distributions are the two sides a thought to be charged with linear density λ such that $\lambda/\epsilon_0 = \mu_0 I$, one positive and one negative (corresponding to the sign of j_x). Considering that the distance of P is large, we can assume the two charges to be point-like. The potential is then the electric potential of a dipole of moment \mathbf{p} of magnitude $p = \lambda ab$ and direction opposite to the y -axis (from the negative to the positive charge). We have

$$\phi(P) = \frac{1}{4\pi\epsilon_0} \frac{\mathbf{p} \cdot \mathbf{r}}{r^3} = -\frac{\lambda ab}{4\pi\epsilon_0 r^3} \frac{y}{r}.$$

We obtain the x -component of the vector potential with the substitution $\lambda \rightarrow \epsilon_0 \mu_0 I$. We have

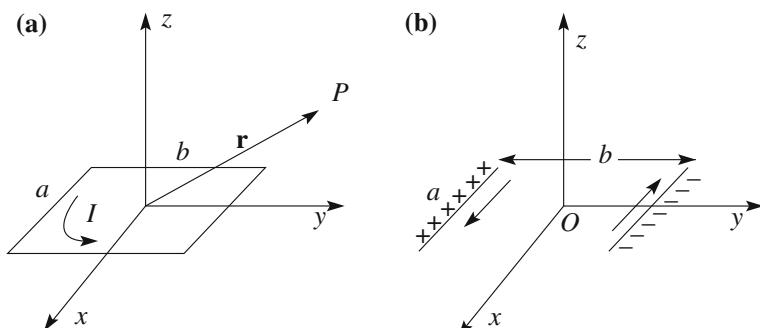


Fig. 6.34 a A magnetic dipole, b its electric equivalent for the x -component of the potential

$$A_x(P) = -\frac{\mu_0}{4\pi} Iab \frac{y}{r^3}. \quad (6.58)$$

A completely analogous calculation for the y -component gives us

$$A_y(P) = \frac{\mu_0}{4\pi} Iab \frac{x}{r^3}. \quad (6.59)$$

The z -component of the vector potential is obviously zero, because there are no currents in the z -direction.

In conclusion, the lines of the vector potential \mathbf{A} are planar curves on the plane of the dipole. The component A_x being proportional to $-y$ and the component A_y proportional to x with the same proportionality constant, these curves are circles with centers in the center of the dipole. \mathbf{A} has the same direction as the current. The magnitude of \mathbf{A} is Iab , namely the product of the current intensity and the area of the loop. As we shall soon see, this statement holds independently of the shape of the loop. We define as the magnetic moment of the dipole the axial vector $\boldsymbol{\mu}$ having the direction normal to the loop, oriented to see the current circulating in a counter-clockwise direction and having a magnitude equal to the current intensity times the area of the loop. With this definition and the components we found, the vector potential of the dipole is

$$\mathbf{A} = \frac{\mu_0}{4\pi} \frac{\boldsymbol{\mu} \times \mathbf{r}}{r^3}. \quad (6.60)$$

To be precise, note that the magnetic moment is an axial vector, rather than a vector, namely its Cartesian components do not change sign under inversion of the reference frame axes. This property is a consequence of the “anticlockwise” specification in its definition. We reach the same conclusion by looking at Eq. (6.60), knowing that \mathbf{A} and \mathbf{r} are proper vectors and that the vector product of a vector and an axial vector is a proper vector.

Let us now find the magnetic field of the dipole. We just need to take the curl of \mathbf{A} , namely

$$\begin{aligned} B_x &= \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} = 0 - \frac{\mu_0}{4\pi} \mu \frac{\partial}{\partial z} \frac{x}{r^3} = \frac{\mu_0}{4\pi} \mu \frac{3xz}{r^5} \\ B_y &= \frac{\partial A_x}{\partial z} = \frac{\mu_0}{4\pi} \mu \frac{\partial}{\partial z} \left(-\frac{y}{r^3} \right) = \frac{\mu_0}{4\pi} \mu \frac{3yz}{r^5} \\ B_z &= \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} = -\frac{\mu_0}{4\pi} \mu \left(\frac{1}{r^3} - \frac{3z^2}{r^5} \right) \end{aligned} \quad (6.61)$$

Comparing these equations with Eqs. (1.92) and (1.93), we see that the \mathbf{B} field of the magnetic dipole has exactly the same dependence on the coordinates as the \mathbf{E} field of the electric dipole, but, remember, at large distances from the dipole.

At first sight, it appears very surprising that two sources as different as a small loop and two point charges governed by equations as different as $\nabla \cdot \mathbf{E} = \rho/\epsilon_0$ and $\nabla \times \mathbf{B} = \mu_0 \mathbf{j}$ give origin to fields of equal shape. However, this is true only at a distance from the sources, namely in the empty space, where both \mathbf{E} and \mathbf{B} have curl and divergence equal to zero. Contrastingly, near the sources, where the right hand sides of the above expressions are relevant, the fields are completely different, as shown in Fig. 6.35.

We now demonstrate the statement we made that the magnetic field of a planar current loop of any shape is equal to the field of a rectangular loop at a large distance.

Consider a planar loop of arbitrary shape carrying a steady current of intensity I , as shown in Fig. 6.36. We can divide it into a network of small rectangular loops, carrying the same current I . The system is equivalent to the given loop, because all the internal segments we have added carry two equal and opposite currents, which cancel one another out. Each of the rectangular loops gives origin to a vector potential proportional to its dipole moment. Let $\boldsymbol{\mu}_i$ be the moment of the i -th loop. All these vectors have the same direction, and magnitude equal to the product of the current intensity (which is the same for all) times the area, say ΔS_i . The vector potential is then

$$\mathbf{A} = \sum_i \mathbf{A}_i = \frac{\mu_0}{4\pi} \frac{\sum_i \boldsymbol{\mu}_i \times \mathbf{r}_i}{r^3} = \frac{\mu_0}{4\pi} \frac{(\sum_i \boldsymbol{\mu}_i) \times \mathbf{r}}{r^3}$$

where, on the right-hand side, we have considered all the position vectors \mathbf{r}_i from a loop to the point P in which we evaluate the vector potential to be equal (to \mathbf{r}), considering that the distance of P is large. Now, we have, for the magnitude, $\sum_i \mu_i = I \sum_i \Delta S_i = IS$, where S is the total area of the loop. Hence, in vector form, $\sum_i \boldsymbol{\mu}_i = \boldsymbol{\mu}$, where $\boldsymbol{\mu}$ is the magnetic moment of the loop.

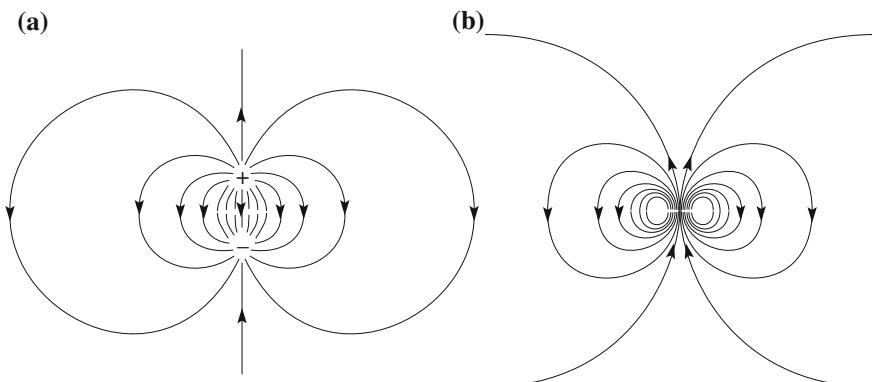
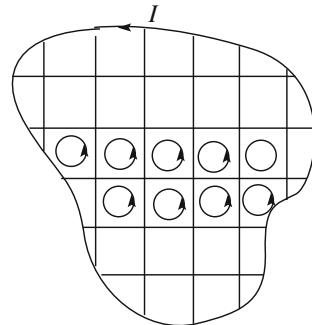


Fig. 6.35 **a** Electric field of an electric dipole, **b** magnetic field of a magnetic dipole

Fig. 6.36 A planar current loop of arbitrary shape



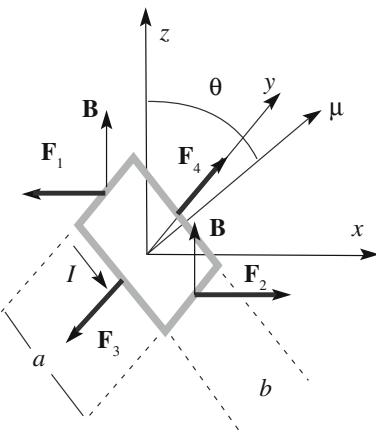
Let us go back for a moment to the circular loop we considered in Sect. 6.11, where we expressed the magnetic field at the points of the axis. We found \mathbf{B} to be directed the same as the axis and given by Eq. (6.55). There, the distance from the center is called z . Let us indicate it with r so as to have the same symbol as in the present discussion. If the distance is large, namely if $r \gg R$, we can approximate the denominator as $(R^2 + r^2)^{3/2} \approx r^3$. We also notice that the magnetic moment of the loop is $\mu = \pi R^2 I$, and Eq. (6.55) becomes

$$B_z = \frac{\mu_0}{2\pi} I \frac{R^2}{r^3} = \frac{\mu_0 \mu}{2\pi r^3},$$

which coincides with Eq. (6.61) evaluated on the axis, namely for $x = y = 0$ and $z = r$.

Let us consider now the actions on a magnetic dipole of an external magnetic field \mathbf{B} , which we shall consider, for the sake of simplicity, uniform over the dimensions of the dipole. We consider a rectangular loop and choose the reference frame, as shown in Fig. 6.37, with the z -axis in the direction of \mathbf{B} and the x -axis in the plane defined by \mathbf{B} and $\boldsymbol{\mu}$. Let θ be the angle between \mathbf{B} and $\boldsymbol{\mu}$.

Fig. 6.37 The forces on a magnetic dipole in a uniform magnetic field



Let us find the torque acting on the dipole. As shown in the figure, four forces act on the loop, one on each of its sides. Two of these, \mathbf{F}_3 and \mathbf{F}_4 , make a couple of zero arms and consequently with zero torque. The other two, \mathbf{F}_1 and \mathbf{F}_2 , have the same magnitude $F_1 = F_2 = IBb$ and the opposite direction. They make a couple of arms $a \sin \theta$, having its torque in the negative y direction. The components of the torque are $\tau_x = \tau_z = 0$ and $\tau_y = -IBba \sin \theta = -\mu B \sin \theta$. In vector form, this is

$$\boldsymbol{\tau} = \boldsymbol{\mu} \times \mathbf{B}. \quad (6.62)$$

It can be that this result found for a rectangular loop is valid for any shape, provided the loop is small enough to have negligible variations of the field on its dimensions. The analogy with the electric dipole is very close. With the same argument we made in that case, one finds that the energy of the magnetic dipole in an external field \mathbf{B} is

$$U_{\text{mec}} = -\boldsymbol{\mu} \cdot \mathbf{B}. \quad (6.63)$$

However, there is an important difference. Equation (6.63) takes into account only the mechanical work needed to put the dipole in position. This is the reason for the subscript. Equation (6.63) does not include the electrical work that might be necessary to maintain the current as constant. We shall come back to this issue in Sect. 8.3.

QUESTION Q 6.6. A square loop of 5 mm side, carrying a steady current of 5 A is in a uniform magnetic field $B = 0.5$ T. Its normal is at 30° with the field direction. What is the mechanical energy of the loop? What is the torque on the loop? \square

6.14 Charge Densities in a Current-Carrying Wire

As we know, the charge density and the electric field inside a conductor, say a metal, in electrostatic equilibrium are zero. Contrastingly, both volume and surface charges are usually present on a current-carrying conductor, a metal wire, for example. In addition, both an electric and a magnetic field exist. We shall consider here steady currents and look at a few examples as to how charge distributions and fields depend on the geometry of the problem. Let us work in the reference frame in which the wire is at rest, which we call the laboratory frame. We can think of the metal as being a crystalline structure of ions. We shall assume the ions to be singly charged, at rest in fixed positions. The wire also contains an equal number of free electrons, which are the charge carriers. Let \mathbf{v}_d be their drift velocity and n_p their numerical density. The current density is then

$$\mathbf{j} = -n_p q_e \mathbf{v}_d, \quad (6.64)$$

where $-q_e$ is the electron charge. In the laboratory frame, the wire is globally neutral, namely its net charge *per unit length* is zero. However, both volume and surface charge densities exist, as we shall now discuss.

Let us start from the simplest case of a straight wire. To produce a steady current density, as in Eq. (6.64), we need, for Ohm's law, to have the electric field \mathbf{E} inside the wire:

$$\mathbf{E} = \rho_{re}\mathbf{j}, \quad (6.65)$$

where we have called the resistivity ρ_{re} to avoid any confusion with the charge density. Equation (6.65) tells us that, under steady conditions, the lines of electric field inside the conductor are parallel to the current lines, which, in turn, are parallel to the wire itself, because there is no current through the lateral surface. Let us see how the field is established. When we connect the wire to the emf source, for example, to the poles of a battery, the field initially extends well beyond the surface of the wire with a number of field lines crossing the surface. The field configuration is "corrected" by a (small) fraction of the conduction electrons that immediately move on the surface. The resulting net surface charge density guides the field lines to be parallel to the wire, preventing current flow in directions different from the wire axis.

This happens for any shape of the wire, including when it is straight. If the wire is bent, as in Fig. 6.38, additional surface charges are necessary. Indeed, the field must be uniform in magnitude and have a direction following that of the wire, because the current density is such. As the electrons move from A to B , they should turn near C in the direction CD , namely to their right. They are forced to do so by a negative surface charge present on their left, which repels them, and a positive charge on their right, which attracts them, as shown qualitatively in Fig. 6.38. Similarly, near D , they must turn to the left, and the surface charges are as shown in Fig. 6.38.

The surface charge densities are necessary, but quantitatively very small. To appreciate the point, let us consider the extreme situation of a sharp 90° bend, as shown in Fig. 6.39, where we have chosen the x and y axes parallel to the two segments of the wire. The figure shows separately the lines of the x and y components of the electric field, namely \mathbf{E}_x and \mathbf{E}_y (clearly, the lines of the total field \mathbf{E} do not cross one another, but curve following the bend).

Fig. 6.38 Surface charges on a bent current-carrying metal wire

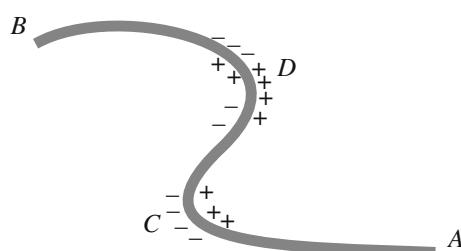
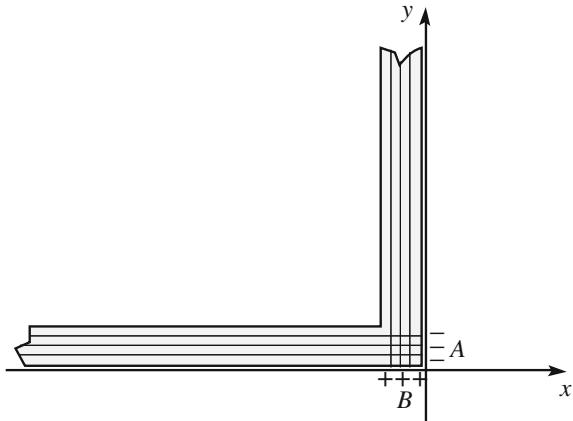


Fig. 6.39 Surface charges on a 90° bend of a current-carrying metal wire. The lines of the \mathbf{E}_x and \mathbf{E}_y components of the electric field are shown



Let Σ be the cross-section of the wire and E the magnitude of the electric field. The \mathbf{E}_x lines from the left must end at the extreme of the straight segment on the right at A . For the Gauss theorem, a charge Q_- must be present in A equal to ϵ_0 times the flux of E , which is ΣE , namely

$$Q_- = -\epsilon_0 \Sigma E = -\epsilon_0 \Sigma j \rho_{re} = -\epsilon_0 I \rho_{re}, \quad (6.66)$$

where we have used Ohm's law [Eq. (6.65)]. Similarly, a surface charge $Q_+ = -Q_-$ should be present at B to give origin to the \mathbf{E}_y field component in the perpendicular segment of the wire. Considering the example of a copper wire, with $\rho_{re} = 1.56 \times 10^{-8} \Omega \text{ m}$, we have $Q_+ = -Q_- = 1.38 \times 10^{-19} I$ [C]. We see that the charge of just about one electron is enough to turn the current of one ampere by 90°. Note that we are talking here of a macroscopic charge, meaning a charge averaged on macroscopic dimensions. This average charge is not quantized and can have any value, even one smaller than the elementary charge. For example, the just made conclusion shows that Q_+ is about one thousandth of the elementary charge for a 1 mA current.

We now go back to the straight metal wire and consider a further effect. The electric current generates a magnetic field \mathbf{B} not only outside the wire but inside it as well. The lines of \mathbf{B} are circles normal to the wire, centered on its axis. The symmetry of the problem imposes that the magnitude of B , the magnitude j of the current density and the volume charge density ρ can be functions only of the distance from the axis, which we call r , and not of the azimuth about the axis. The magnitude of the magnetic field $B(r)$ inside the wire, namely for $r < R$, where R is the radius of the wire, is given by the Ampère circulation law. It is

$$B(r) = \frac{\mu_0}{2\pi r} \int_0^r j(r) 2\pi r dr. \quad (6.67)$$

A consequence of this field is that the conduction electrons are subject to the Lorentz force $-q_e \mathbf{v}_d \times \mathbf{B}$, which is directed radially towards the axis. Knowing that the electrons move parallel to the axis, we conclude that an electric field $\mathbf{E}_r = -\mathbf{v}_d \times \mathbf{B}$ must exist to balance the Lorentz force. The effect is similar to the Hall effect. This field is generated by the following mechanism. Initially, the electrons are pushed towards the axis. In moving in that direction, they leave a thin layer near the surface in which an excess of positive charge density appears. This is a charge of the ion ladder. This charge produces an electric field directed radially towards the axis that pushes the free electrons radially outside. The process terminates when the equilibrium between the two forces is reached. The result is the presence of both a surface positive charge density and a volume charge density. The latter is given by the Gauss law,

$$\rho(r) = \epsilon_0 \nabla \cdot \mathbf{E}_r = -\epsilon_0 \nabla \cdot (\mathbf{v}_d \times \mathbf{B}),$$

which can be simplified using a vector identity as

$$\rho(r) = -\epsilon_0 \mathbf{B} \cdot (\nabla \times \mathbf{v}_d) + \epsilon_0 \mathbf{v}_d \cdot \nabla \times \mathbf{B}.$$

But \mathbf{v}_d is uniform, and consequently, $\nabla \times \mathbf{v}_d = 0$. Substituting $\nabla \times \mathbf{B} = \mu_0 \mathbf{j}$, we have

$$\rho(r) = \mu_0 \epsilon_0 \mathbf{v}_d \cdot \mathbf{j} \quad (6.68)$$

Here, we meet for the first time the product $\epsilon_0 \mu_0$ of the basic constants of electrostatics and magnetostatics. A dimensional analysis shows that its physical dimensions are those of an inverse velocity squared. We shall see in Chap. 10 how James Clerk Maxwell discovered that velocity should be the speed of light, c . Also taking into account Eq. (6.64) and recalling that the current density \mathbf{j} is parallel and opposite to the electron drift velocity \mathbf{v}_d , we can eliminate the current density from Eq. (6.68) and write

$$\rho(r) = -n_p q_e \frac{v_d^2}{c^2} = -n_p q_e \beta_d^2 \quad (6.69)$$

where $\beta_d = v_d/c$ is the ratio between drift velocity and speed of light. We see that the volume charge density does not depend on the distance from the axis r . Let us consider as an example of a copper wire of $\Sigma = 1 \text{ mm}^2$ cross-section carrying a current of intensity $I = 1 \text{ A}$, corresponding to a current density $j = 10^6 \text{ A m}^{-2}$. In a copper crystal, the ions are singly ionized. Consequently, the numerical densities of ions and conduction electrons are equal. Their value is $n_p = 8.47 \times 10^{28} \text{ m}^{-3}$. The drift velocity is $v_d = j/(n_p q_e) = 7.4 \times 10^{-5} \text{ m s}^{-1}$, corresponding to $\beta_d = 2.47 \times 10^{-13}$. The charge density is then

$$\begin{aligned}\rho &= -n_p \beta_d^2 q_e = -q_e 8.47 \times 10^{28} \times (2.47 \times 10^{-13})^2 = -q_e 5.17 \times 10^3 \text{ m}^{-3} \\ &= -3.23 \times 10^{-16} \text{ Cm}^{-3},\end{aligned}$$

namely an average charge of 5.000 electrons per cubic meter. This is a very small charge indeed.

The (negative) charge density Eq. (6.69) is due, as we discussed, to the contraction of the conduction electrons toward the axis of the wire, under the action of the magnetic force. In doing so, the electrons leave a very thin layer of positive ion net charge. This surface positive charge density, which we call σ , makes the total charge per unit length of the wire equal to zero. If R is the radius of the wire, this condition is $2\pi R\sigma = -\pi R^2 \rho$, which gives us

$$\sigma = -\rho R/2 = -q_e n_p \beta_d^2 R/2. \quad (6.70)$$

This is an extremely small density, even on the atomic scale. Indeed, in the above example, with a wire cross-section $\Sigma = 1 \text{ mm}^2$, namely a radius $R = \sqrt{\Sigma/\pi} = 0.56 \text{ mm}$, and $\rho = -3.23 \times 10^{-16} \text{ Cm}^{-3}$, we have $\sigma \cong 0.56 |q_e| \text{ m}^{-2}$. Let us consider for comparison just one layer of ions. The average numerical surface ion density is equal to their volume number density quoted above, $n_p = 8.47 \times 10^{28} \text{ m}^{-3}$, times the size of the crystal cell, which is $a = 0.361 \text{ nm}$, namely about $3 \times 10^{19} \text{ m}^{-2}$. The total surface charge density of a single ionic layer is then $\sigma_{\text{ion}} = 3 \times 10^{19} |q_e| \text{ m}^{-2}$. This is about 5×10^{19} times larger than σ , implying that the ion sheet whose charge is not compensated by electrons is 2×10^{-20} thinner than an atomic layer. Indeed, cases exist in nature in which very small numbers are sufficient to reach a scope.

6.15 Properties of the Laws Under Rotations and Inversions of the Axes

In this section, we shall discuss an important property of the mathematical expressions of the physical laws, namely how they transform under a change of reference frame. We shall repeat and extend to electromagnetism concepts already discussed in the 1st volume in Sects. 1.3 and 5.1. We shall first consider the transformations between two frames at rest relative to one another. The translations are not very interesting, because they merely correspond to a change of the origin. The interesting transformations are the rotations and the inversions of the axes. All the experiments have shown that the laws of physics are covariant under the rotations of the axes. “Covariance” means that the two sides of an equation, if they vary, should vary in the same way. This means that if the left-hand side is, for example, a vector, the right-hand side must be a vector too; if it is a scalar, the right-hand side must also be a scalar.

We recall that a quantity is a scalar if it does not vary when we change reference frame. A vector is an ordered triplet of real numbers, which transform under rotations of the axes as the position vector (namely as the coordinates, which are the components of the position vector).

Consider, for example, the relation between two scalar quantities:

$$a = b. \quad (6.71)$$

If it holds in a reference frame, it also holds in another frame linked to the first by a rotation, because neither the left-hand side nor the right-hand side vary ($a' = a$, $b' = b$). Contrastingly, the relation

$$a = u_x \quad (6.72)$$

where a is a scalar and u_x is the x -component of a vector, is not covariant, namely $a' = a$ and $u'_x \neq u_x$. Hence, if the relation holds in the first reference, it does not hold in the second. Note that in order for a relation to be valid in both frames, it is not necessary that its members do not vary, rather they should vary in the same way. This is the case, for example, in the relations between analogous components of vectors like

$$v_x = u_x, \quad v_y = u_y, \quad v_z = u_z \quad (6.73)$$

This is the reason for the word covariance, meaning to vary together. We state that the invariance of the physical laws under rotations implies the covariance of the equations expressing those laws.

Let us now consider the inversions of the axes. We can invert one, two or three axes. However, it is immediately seen that the inversion of two axes is equivalent to a rotation by 180° about the third axis. In addition, the inversion of one axis and that of three axes are linked by a rotation. However, the inversion of one axis (and of three) cannot be obtained with rotations.

By definition, a proper scalar quantity is invariant both under rotations *and* under inversion of the axes. A pseudoscalar quantity is invariant under rotations and changes sign under inversion of the axes. A proper vector quantity has three components that change in the same way as the position vector, both under rotations *and* under inversion. Namely, under inversion of the axes, the components of a vector change sign. A pseudovector, also called an axial vector, transforms in the same way as the position vector under rotations but its components do not change sign under inversion. Examples of vectors are the velocity, the acceleration, the linear momentum and the force. As a consequence, the electric field is a vector, being a force divided by a charge, which is a scalar.

The cross-product of two vectors is an axial vector, because when the axes are inverted, both factors change sign and the sign of the product does not. Examples of axial vectors are the angular velocity, the moment of a force, the angular momentum and the magnetic dipole moment.

The axial vector nature of a quantity becomes evident considering that its definition always includes terms such as “clockwise” or “right-hand”. The right hand becomes left hand and the clockwise direction becomes counter-clockwise, inverting one axis.

The cross-product of two axial vectors is a (proper) vector. The scalar product of two vectors and of two axial vectors is a scalar. The scalar product of a vector and an axial vector is a pseudoscalar. In particular, the scalar triple product of three vectors is a pseudoscalar.

The laws of electromagnetism are invariant not only under rotations but also under inversion of the axes. Let us analyze the behavior under inversions of the magnetic field. The definition of the direction of the magnetic field includes the right hand rule or terms like ‘counter-clockwise direction’. This is implicit in expressions such as the 1st Laplace law

$$d\mathbf{B}(\mathbf{r}_2) = \frac{\mu_0}{4\pi} I \frac{d\mathbf{s} \times \mathbf{u}_{12}}{r_{12}^2}. \quad (6.74)$$

and of the definition of the cross-product in a right-handed reference frame. At first sight, this might suggest that electromagnetic phenomena distinguish between left and right, or, in other words, are not invariant under inversion of the axes (parity operation). This is not true, however, because all the physical observables result from applying the right-hand rule twice, and inverting axes twice is like doing nothing. Indeed, the magnetic field is observable only through the force it exerts on moving charges, which is the Lorentz force that involves a (second) cross-product.

Two parallel currents attract one another if they have the same direction, and repel if their directions are opposite, whatever hand-rule we choose.

As another example, the electric dipole moment \mathbf{p} is a vector, while the magnetic dipole moment $\boldsymbol{\mu}$ is an axial vector. The energy of the electric dipole $\mathbf{p} \cdot \mathbf{E}$ and of the (mechanic) energy of the magnetic dipole $\boldsymbol{\mu} \cdot \mathbf{B}$ are both scalar, the former as the dot product of two vectors, the latter as the dot product of two axial vectors.

In complete generality, we can state that all electromagnetic phenomena are independent of the handiness of the reference frame. This basic symmetry is a characteristic of both electromagnetic and gravitational interactions. At the microscopic level, it is a characteristic of the strong nuclear interaction (which keeps protons and neutrons linked in the nuclei) as well, but not of the weak interaction (responsible for the beta decay). The invariance under inversion of the axes is *not* a universal law of Nature.

6.16 Relativity of Electric and Magnetic Forces

The relativity principle established by Galileo Galilei (Italy, 1564–1642) states that physical laws are invariant under transformations between two inertial reference frames in relative motion of uniform translation. As discussed in the 1st volume,

only two sets of transformations of the space and time coordinates between such frames exist (under very general assumptions): the Galilei transformations and the Lorentz transformations. The former are the limit of the latter for velocities much smaller than the speed of light. In the Galilei transformations, time intervals and distances between two points are invariant under the transformation, namely they are the same for two observers in relative motion. Contrastingly, they are not the same under Lorentz transformations, which mix time and space. As we shall see in Chap. 10, the relativity principle holds for the laws of electromagnetism, provided that space and time coordinates transform according to the Lorentz transformations. Let us analyze the behavior of the electric and magnetic force now.

The force acting on a point charge q moving with velocity \mathbf{v} is

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \quad (6.75)$$

Clearly, the magnetic force $q\mathbf{v} \times \mathbf{B}$ is different in two frames in relative motion, because the velocity of the particle is different. In particular, the magnetic force is zero in the frame in which the charge is instantaneously at rest. However, if the relativity principle has to be satisfied, the total force, electric plus magnetic, should transform, as required by the Lorentz transformations. This implies that electric and magnetic fields can exchange their roles when the velocity of the reference frame changes. In other words, the relativity principle imposes that electric and magnetic fields be part of a unique entity. Let us discuss the issue.

We first recall that the electric charge is invariant under Lorentz transformations. We shall now see that the current density \mathbf{j} , which is a three-vector (namely a vector in the three-dimensional space), and the charge density ρ , which is a three-scalar (namely a scalar in the three-dimensional space), when multiplied by suitable constants, form a four-vector (namely a vector in the four-dimensional space-time). The four-vector is $(\mathbf{j}, i\rho c)$. We shall prove this statement using the analogy with the energy-momentum four vector $(\mathbf{p}, iU/c)$ of a point-like particle. We recall that the momentum of a particle of mass m is $\mathbf{p} = m\gamma\mathbf{v}$ and the energy is $U = mc^2$, where $\gamma = 1/\sqrt{1 - \beta^2}$ and $\beta = v/c$, as we learned in Chap. 6 of the 1st volume.

Let \mathbf{j} and ρ be, respectively, the current and charge density in the inertial reference frame S . Let us assume for simplicity that all the charge carriers have the same (drift) velocity \mathbf{v}_d . The current density is then $\mathbf{j} = \rho \mathbf{v}_d$. Let S_0 be the reference frame in which the charge carriers are at rest. S_0 translates relative to S with velocity \mathbf{v}_d . Being that \mathbf{v}_d is constant, S_0 is inertial too. We take the x -axis of S and x' -axis of S_0 , both parallel to \mathbf{v}_d . We indicate with the suffix “ $_0$ ” the quantities in S_0 . Consider a certain charge ΔQ of the current, which is at rest in S_0 (no suffix on it, because the charge is invariant). If ΔV_0 is the volume taken by ΔQ , the charge density (at rest) is $\rho_0 = \Delta Q/\Delta V_0$. The current density is clearly $\mathbf{j}_0 = 0$. Going now to S , the charge does not vary, but the volume does. The dimensions perpendicular to the translation, namely to \mathbf{v}_d , do not vary, but the longitudinal dimension contracts by the factor $1/\gamma_d$,

where $\gamma_d = 1/\sqrt{1 - \beta_d^2}$ and $\beta_d = v_d/c$. The volume in S is $\Delta V = \Delta V_0/\gamma_d$ and the charge density is consequently $\rho = \gamma_d \rho_0$. We then have

$$\mathbf{j} = \rho_0 \gamma_d \mathbf{v}_d, \quad \rho = \rho_0 \gamma_d. \quad (6.76)$$

These quantities can be expressed in terms of the components of the four-momentum of a particle of velocity \mathbf{v}_d , which we call $(\mathbf{p}_d, iU_d/c)$, as

$$\mathbf{j} = \frac{\rho_0}{m} \mathbf{p}_d, \quad \rho c = \frac{\rho_0}{m} \frac{U_d}{c}. \quad (6.77)$$

Hence, the components of $(\mathbf{j}, i\rho c)$ are proportional to those of $(\mathbf{p}_d, iU_d/c)$ with the proportionality constant ρ_0/m that is independent of the frame. As a consequence, $(\mathbf{j}, i\rho c)$ is a four-vector that transforms between two inertial reference frames S and S' moving with relative velocity \mathbf{v} as

$$j'_x = \gamma(j_x - \beta \rho c); j'_y = j_y; j'_z = j_z; \rho' c = \gamma(\rho c - \beta j_x) \quad (6.78)$$

where $\gamma = 1/\sqrt{1 - \beta^2}$ is the Lorentz factor and $\beta = v/c$.

We now go back to the Lorentz force. We consider a straight wire of section Σ carrying a steady current of intensity I in an inertial frame S that we shall call the laboratory frame. We take the x -axis of S parallel to the wire in the direction opposite to the current. Consider a point charge q moving with constant velocity \mathbf{v} parallel and opposite to the current at the distance r from the wire, as in Fig. 6.40a. The current generates in the position of the charge the magnetic field of magnitude

$$B = \frac{\mu_0}{4\pi} \frac{2I}{r}$$

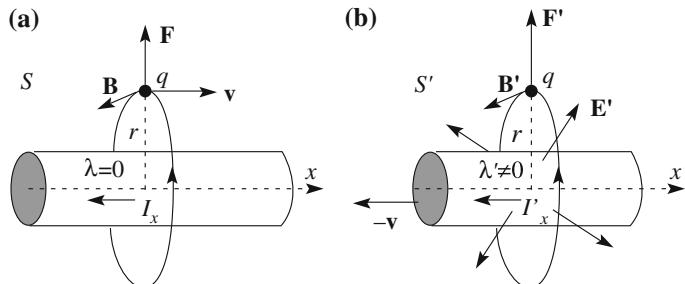


Fig. 6.40 A point charge near a current-carrying wire element, **a** in the frame S in which the wire is at rest; **b** in the frame S' in which the charge is at rest

perpendicular to the wire and consequently to \mathbf{v} . The force on q is purely magnetic, because the wire contains equal numbers of positive and negative charges. The force is directed normal away from the wire. Its magnitude is

$$F = qBv = \frac{\mu_0}{4\pi} \frac{2Iqv}{r}.$$

We do not specify, for the moment, the physical nature of the conductor, namely the nature of the charge carriers, which may be only negative, as in a metal, or both negative and positive, as in a semiconductor or an electrolyte. Let $\rho^-(r)$ and $\rho^+(r)$ be the charge densities of the negative and positive charges in the conductor, respectively. Both densities can be functions of the distance r from the axis of the wire only. They cannot depend on the azimuth around the axis, due to the symmetry of the problem. They are independent of x , because the current is steady. We discussed the case of the metal conductor in Sect. 6.14, but we do not really have any need to know these functions, which depend on the nature of the conductor, because the relevant quantities are the linear charge densities, λ^- for the negative and λ^+ for the positive charges. The linear densities are obtained integrating the volume densities on the section of the wire, namely

$$\lambda^- = \int_{\Sigma} \rho^-(r) d\Sigma; \quad \lambda^+ = \int_{\Sigma} \rho^+(r) d\Sigma. \quad (6.79)$$

Being that any segment of our straight wire is globally neutral, the positive and negative linear charge densities are equal and opposite, namely

$$\lambda^- = -\lambda^+. \quad (6.80)$$

Let us also consider the current densities, which are in the x -direction. Let $j_x^-(r)$ and $j_x^+(r)$ be the current densities due to the negative and positive carriers, respectively. The current densities are functions of the distance from the axis only for the same reasons as the charge densities. By integrating on the section of the wire, we obtain the current intensities due to the positive and negative carriers, namely

$$I^- = \int_{\Sigma} j_x^-(r) d\Sigma; \quad I^+ = \int_{\Sigma} j_x^+(r) d\Sigma. \quad (6.81)$$

Consider now a frame S' translating relative to S with velocity \mathbf{v} . For an observer in this frame, the charge is at rest and the Lorentz force does not act. Does an electric force exist? Let us now consider the conductor to be metallic. In the laboratory frame, the positive charges, which are the ions, are at rest, while the conduction electrons are free to move. They are the charge carriers, and we assume them, for simplicity, all to have the same velocity \mathbf{v}_d , which is directed in the

positive x direction. The wire is globally neutral and Eq. (6.80) holds. The positive and negative current intensities are, in this case,

$$I^- = \lambda^- v_d; \quad I^+ = 0. \quad (6.82)$$

The force on the charge q , which is purely magnetic, is

$$F = qBv = \frac{\mu_0 q}{2\pi} \frac{\lambda^+ v_d v}{r}. \quad (6.83)$$

Let us now consider the situation in S' , as shown in Fig. 6.40b. An observer in this frame sees the point charge q at rest and the wire moving with velocity $-v$. We need to transform the current intensities and the linear charge densities from the laboratory frame S to S' . These quantities have been obtained integrating the x -component of the current density and the volume charge density on the section of the wire, namely on the coordinates normal to the relative velocities of the two frames. Being that these coordinates are equal in the two frames, the transformation equations are not changed by the integration, and from Eq. (6.78), we have

$$I'_x = \gamma(I_x - \beta\lambda c); \quad \lambda'c = \gamma(\lambda c - \beta I_x). \quad (6.84)$$

In S , the positive and negative linear charge densities are related by Eq. (6.80) and the current intensities given by Eq. (6.82). Using the second Eq. (6.84), in S' , the linear charge densities are $\lambda'^- = -\gamma\lambda^+(1 - \beta v_d/c)$ and $\lambda'^+ = \gamma\lambda^+$, which are now different. The total linear charge density in S' is

$$\lambda' = \lambda'^- + \lambda'^+ = \gamma\lambda^+ \frac{v_d v}{c^2}. \quad (6.85)$$

To the observer S' , the wire appears to be charged as a consequence of the Lorentz contraction of the lengths. The consequence is the presence in S' of an electric field. This is the field of a uniformly charged straight wire of section Σ and linear charge density, as in Eq. (6.85). The magnitude of the electric field is

$$E' = \frac{\lambda'}{2\pi\epsilon_0 r} = \frac{\gamma\lambda^+}{2\pi\epsilon_0 r} \frac{v_d v}{c^2}.$$

The currents produce a magnetic field in S' as well. We do not need its expression, because it does not generate any force on the charge, which is at rest. In conclusion, in S' , the charge q is subject to a force, which is purely electric, is directed normally away from the wire, and has the magnitude

$$F' = qE' = \gamma \frac{q}{2\pi\epsilon_0} \frac{\lambda^+}{r} \frac{v_d v}{c^2}. \quad (6.86)$$

Let us now compare the force in S' and in S , given by Eq. (6.83). We see that several factors are equal. However, in Eq. (6.83), we have the characteristic constant of magnetism μ_0 , in Eq. (6.86), the constant of electricity ϵ_0 and the speed of light c . As we have anticipated in Sect. 6.14 and as we shall see in Chap. 10, the light itself is an electromagnetic phenomenon, as discovered by James Clerk Maxwell. His theory, which has been experimentally verified (see Chap. 10), foresees a well-defined value for the speed of light in terms of μ_0 and ϵ_0 , namely

$$c = 1/\sqrt{\epsilon_0 \mu_0}. \quad (6.87)$$

Taking that into account, the relation between the forces in the two frames is

$$F' = \gamma F, \quad (6.88)$$

which is just the correct relativistic relation between the components of the force normal to the relative velocity of the two frames. We can understand this with the following argument.

In the frame S' , the charge q is at rest. The force \mathbf{F} (which is perpendicular to \mathbf{v}), acting on the charge q for a time dt , changes its momentum by $d\mathbf{p}_n = \mathbf{F} dt$ (the footer n recalls that the change of momentum is normal to velocity). From the point of view of S' , the force \mathbf{F}' , acting in the corresponding time interval dt' , changes the momentum of the charge by $d\mathbf{p}'_n = \mathbf{F}' dt'$. Now, $d\mathbf{p}_n$ and $d\mathbf{p}'_n$ are components of a four-vector that are normal to the velocity of the transformation. Consequently, they are equal. The relation between dt and dt' is immediately found, considering that dt' is a proper time interval (namely the time interval between two events happening at the same point). Hence, the relation is $dt = \gamma dt'$. This leads to Eq. (6.88).

We have verified that the force given by Eq. (6.75) obeys the relativity principle if the Lorentz transformations are used. As a matter of fact, observer S and S' observe the same physical process, with the difference that the former interprets the force as being due to the magnetic field, while the latter as being due to the electric field. In a third reference frame, we would find different contributions to the force of both fields. This means that electric and magnetic forces are two aspects of the very same phenomenon. The separation of the *electromagnetic* interaction in electric and magnetic components depends on the reference frame. The complete description, namely the electromagnetic one, is frame independent.

Before concluding this section, we make one further observation. The force acting on the charge, as seen by S' in Eq. (6.86), is proportional to the total charge density, which is, in turn, proportional to the ratio of the product of two velocities, v_d and v , and the square of the velocity of light. The two velocities are extremely small compared to the speed of light. For an evaluation of the orders of magnitude, let us consider what we found in Sect. 6.14 to be the typical value of the drift velocity, namely $\beta_d = v_d/c = 2.47 \times 10^{-13} \text{ m s}^{-1}$. Let v be equal and, say, be 3 m/s, namely $\beta = 10^{-8}$. The force is proportional to their product, which, in order of magnitude, is 10^{-21} . Terms proportional to β^2 , called relativistic effects, are usually completely negligible in everyday phenomena. Looking at the magnetism

from this point of view, we see that it is a relativistic effect proportional to β^2 . For what reason, then, does the magnetism also appear so evident at speeds much smaller than the speed of light? The magnetic “relativistic correction” is a correction of the electric force, which is due to the ions and the electrons. Both contributions are enormous in absolute value and of opposite sign, but the charges of the electron and the proton are exactly equal and opposite, and in matter, there is exactly the same number of protons and electrons. The two huge forces exactly cancel one another out in S , exactly relativistic correction apart in S' . Magnetism is, indeed, a relativistic correction, a correction to a term that is zero due to the perfect balance between positive and negative charges in matter.

Another “relativistic correction” that also appears at everyday velocities is the phenomenon considered in Sect. 6.14, where we found the charge density in a metal current-carrying wire to be proportional to β_d^2 .

Summary

This chapter has two parts. In the first part, we have studied the motion of charged particles in a given stationary magnetic field. We learned the following principal concepts:

1. The magnetic force acting on a charge and how it depends on the motion of the charge.
2. That the force changes the direction of the moment of the charge, but not its energy.
3. How the magnetic field can be measured (Hall effect).
4. The invariance of the electric charge.
5. The exact equality of the elementary positive and negative charges.
6. The motion of point charges in a magnetic field.
7. How to measure current intensities (galvanometer).

In the second part of the chapter, we studied the relations between a magnetic field and its sources, which are the electric currents, limited to the case of steady currents. We learned the following principal concepts:

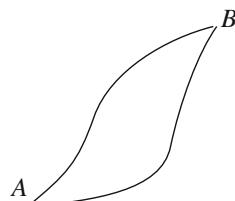
1. The non-existence of magnetic charges. The property of field \mathbf{B} to be a solenoidal field.
2. The relation between currents and the magnetic field they produce.
3. The local relation between spatial derivatives of the magnetic field and current density.
4. The discontinuity of the magnetic field across a current sheet.
5. The vector potential.
6. How to calculate the vector potential produced by a given system of currents.
7. How to calculate the magnetic field produced by a given system of currents.

8. The magnetic dipole. The magnetic moment. Analogies and differences between magnetic and electric dipoles.
9. The current density and charge density as the component of a four-vector.
10. The transformation of magnetic and electric forces one into the other when changing from one reference frame to another in relative motion.

Problems

- 6.1. Can closed lines exist in a stationary magnetic field? And lines coming out of a point?
- 6.2. A He nucleus has the charge $q = 3.2 \times 10^{-19}$ C in its rest frame. What is this charge in a reference frame in which it moves at 90 % of the speed of light?
- 6.3. A point-charge in a vacuum is at 1 mm from the north pole of a permanent magnet. What is the force on the charge?
- 6.4. A straight wire carries a steady current of intensity I . At a distance of 0.5 m from the wire, the magnetic field has the magnitude $B = 3$ mT. What is the value of I ?
- 6.5. With a copper wire, we form two equal circles geometrically overlapped electrically in series and having a steady current throughout. Consider two cases. (a) the current direction is the same in both loops; (b) it is opposite. In which cases is the field in the points of the axis higher?
- 6.6. A straight cylindrical wire of radius R carries a steady current of uniform density j in the direction of the axis. Determine the magnetic field as a function of the distance from the axis of the wire (inside and outside).
- 6.7. A proton moves with a velocity 50 % the speed of light on the axis of a rectilinear solenoid that generates a field of 1 T. What is the force on the proton?
- 6.8. What is the physical reason for \mathbf{B} to be solenoidal?
- 6.9. Consider a point charge moving from A to B once on one path, once on the other of Fig. 6.41, in a magnetic field. Are the works of the magnetic force equal or different?
- 6.10. In a Dempster mass spectrometer, a proton and a positron (that is equal to an electron but with a positive charge) are accelerated with the same potential difference and then enter the magnetic field. What is the ratio of the radii of their orbits?

Fig. 6.41 Two paths from A to B in a magnetic field



- 6.11. The Li nucleus has 3 protons and exists in two isotopes of masses 6 and 7. In a Dempster spectrometer, singly charged Li ions are accelerated with a potential difference of 300 V. The magnetic field is 0.5 T. What are the radiiuses of the orbits?
- 6.12. Do conservative non-solenoidal fields exist? Do non-conservative solenoidal fields exist?
- 6.13. An Al foil (density = 2.6×10^3 kg/m³, atomic number = 27), has thickness $c = 0.2$ mm, width a and length b . It carries a steady current of intensity $I = 5$ A through the section ac . The foil is in a magnetic field normal to the face ab of intensity $B = 0.5$ T. Calculate the Hall voltage. (There is one conduction electron per atom.)
- 6.14. A cyclotron has the magnetic field $B = 1.2$ T, radius of the D_s $R = 0.5$ m and accelerating emf $V = 50$ kV. Find the maximum reachable kinetic energy of the protons and the corresponding velocity. How many times do the protons go through the gap between the D_s before reaching the maximum energy? How much time do the protons spend in the cyclotron?
- 6.15. A Fe nucleus of cosmic rays enters the earth's magnetic field with 10 nJ energy. After having been deflected by the field, it exits with 12 nJ energy. Is this possible?
- 6.16. The magnetic flux in a rectilinear solenoid of length $l = 25$ cm is $\Phi_B = 50$ μ Wb. Neglecting the edge effects, what is the magnetic moment of the solenoid?
- 6.17. A square loop of 5 cm side is in a uniform magnetic field $B = 1.2$ T with its normal forming an angle of 30° with the field. What is the magnetic flux through the loop?
- 6.18. A straight wire carries a steady current of intensity $I = 3$ A. What are the direction and the magnitude of the magnetic field at 1 cm from the wire? And at 2 cm?
- 6.19. A coil, of which we do not know the shape but we know its dimensions to be on the order of centimeters, carries a steady current. The field on the axis at 1 m distance from the center is 20 mT. How much is the field on axis at 10 m?
- 6.20. Two identical circular coils are located parallel to one another on the same axis at a distance of 2 m. Their radius is $R = 1$ cm. The current intensity is $I = 100$ A. How much is the interaction energy? How much is the exchanged force?
- 6.21. A source of electrons S emits electrons having velocities equal in magnitude $v = 3 \times 10^6$ m/s and directions slightly divergent around the x -axis. A magnetic field of intensity $B = 10$ mT is directed along the x -axis. Find the distance of the first focus from S .
- 6.22. How can you produce a circular line of \mathbf{B} ? And one of \mathbf{A} ?
- 6.23. An infinite plane current sheet produces a vector potential of 10^{-3} T m. What are the magnitude and the direction of \mathbf{A} at 2 m?

- 6.24. A rectilinear solenoid carries a steady current. Outside the solenoid, at a distance of 1 m from the axis, the vector potential magnitude is A . How much is it at 10 m from the axis?
- 6.25. An electron moving with $v = 2 \times 10^6$ m/s enters a solenoid normally to its axis. The solenoid is made of $n = 5000$ loops/m and carries a current of intensity $I = 20$ A. What is the curvature radius of the electron's trajectory?
- 6.26. We build a cube by assembling six identical copper wires (its dihedral angles). We send a current through two opposite corners. What is the magnetic field in the center? (Hint: think of the symmetry of the problem.)

Chapter 7

Electromagnetic Induction

Abstract In this chapter, we discuss the electromagnetic induction phenomenon, starting from the discovery by Michael Faraday that when the magnetic flux linked to a circuit varies with time, an electromotive force, and its consequent current, appears in the circuit. We learn the differential equation ruling the phenomenon, namely that the curl of the electric field is equal to the opposite of the time derivative of the magnetic field. We discuss Faraday's law under various circumstances and introduce the concepts of mutual inductance and self-inductance. Finally, we deal with the alternate current circuits and the impedances of their different passive components, resistors, capacitors and inductors.

Up to now, we have discussed the electric and magnetic phenomena in time independent situations. Under such conditions, the electric charge density ρ is a function of the space coordinates but not of time, and originates a constant electric field. Similarly, the current density \mathbf{j} does not vary with time, and originates a stationary magnetic field. The differential equations ruling the electric field contain, as a source, the charge density, while those for the magnetic field only the current density. Time independent electric and magnetic fields appear to be two independent quantities. All of this is not true under dynamic conditions, namely when the charge density, the current density and the fields are functions of time. We shall now begin to explore the fact that electric and magnetic fields are not at all separate entities, but rather are so intimately connected so as to be components of a single quantity, the electromagnetic field.

In this chapter, we discuss the electromagnetic induction phenomenon, discovered in the first half of the XIXth century by Michael Faraday (and others). When the magnetic flux linked to a circuit varies with time, an electromotive force, and its consequent current, appears in the circuit. The phenomena are summarized in the flux rule. This rule is, in fact, a double rule. From one side, it is a consequence of the Lorentz force, which we did already know; from another, it is a consequence of a new general law. This is Faraday's law stating how the electric field is produced by a magnetic field varying with time. More precisely, we shall see that the curl of the electric field is equal to the opposite of the time derivative of the magnetic field.

Faraday's law for electromagnetic induction is at the basis of electro technics, a vital component of our civilization. In the second part of the chapter, we discuss a number of (non-technical) applications of Faraday's law. We introduce the concepts of mutual inductance between two circuits and self-inductance of a single circuit on itself.

In the last three sections, we study alternate current circuits and the impedances of their different passive components, resistors, capacitors and inductors.

7.1 The Flux Rule

In this chapter, we shall start our study of the electric and magnetic fields under dynamic, namely time-dependent, conditions. Soon after the discovery by Hans Christian Ørsted in 1820 that an electric current, namely moving electric charges, produce a magnetic field, scientists started to suspect that some connection had to exist between electricity and magnetism. The discovery of the relationship was not easy at all. It took more than 10 years and the careful work of several researchers, above all the genius of Michael Faraday (UK, 1791–1867), but also of Joseph Henry (USA, 1797–1878), Franz Neumann (Germany, 1798–1895) and even more in the subsequent years. The main question was whether a magnetic field produced, for example, by a current in a circuit was able to generate, or to "induce", an electric current in a second circuit. Following this idea, Faraday took an insulating cylinder and wrapped an insulated metal wire about it in the form of a densely packed helix (primary circuit). He wrapped a second insulated wire above the first (secondary circuit). He connected the ends of the secondary to a sensitive galvanometer and then those of the primary to the poles of a pile. Note that he was aiming to work with a continuous current, namely constant in time. In his first experiments, he could not detect any current in the secondary. He developed galvanometers capable of the greatest sensitivity he could obtain; but he still did not observe any effect. However, it happened that he did observe short current pulses just after having closed the switch connecting the primary to the battery, and just after having opened it as well. He thought that perhaps a variable current was necessary.

We observe that Faraday, starting from a failed experiment, had reached an outstanding discovery, taking advantage, with his genius and intuition, of a fortunate observation. This is a historical example of how the progress of our knowledge often follows a tortuous path. Only after the experiments have clarified a phenomenon in all its aspects, through a trial and error process, can the issue be theoretically arranged into a picture that appears simple and elegant.

Starting from his preliminary conclusion, Faraday developed a series of experiments to understand completely the electromagnetic induction phenomenon, as it is called. He found three basic different ways to induce a current in a circuit. Consider a secondary circuit consisting of many loops of a metal wire wrapped around an insulating cylindrical tube and connected in series to a sensitive galvanometer. We can classify the current induction cases in the following categories.

(a) *Moving the circuit in a non-uniform steady magnetic field.*

Let us take, for example, a cylindrical permanent magnet having a diameter smaller than that of the tube supporting the secondary circuit. If we keep the magnet still with one hand and move the circuit with the other one so as to have the magnet penetrating the tube, we observe an induced current, as long as we move the circuit. The current disappears as soon as we stop moving. We also observe that the induced current intensity is larger when we move the circuit faster. If we move the circuit in the opposite direction, the induced current flows in the opposite direction. As a variant, we observe that moving just a segment of a circuit in a magnetic field is enough to induce a current. We can use a magnet and a secondary circuit that includes a movable segment. When we move the segment into the field region, we observe an induced current. When we stop the motion, the current vanishes. When we take the segment out of the field, the induced current has the opposite direction.

(b) *Moving a magnet near to the circuit.*

Using the same secondary circuit as before, we also observe an induced current when, keeping the circuit still, we insert the magnet or extract it. Indeed, the effect depends on the relative motion of the two objects. Observe, however, that the two situations are radically different. In case (a), the circuit moves and the magnetic field is independent of time. In case (b), the circuit is at rest and the magnetic field varies in time at its points. Indeed, moving the magnet closer or farther away is just a way to create a time-dependent magnetic field.

(c) *Having a variable current intensity in another (primary) circuit close by.*

The Faraday experiment from which we started this section falls into this category. We can still use our cylindrical secondary circuit to detect an induced current. We build another circuit, the primary, by wrapping another helix of metal wire around another hollow cylinder having a diameter large enough to allow our secondary circuit to enter. The primary circuit is connected to a battery. Moving the secondary in or out of the primary at rest, we observe an induced current. This is a variant of category (a). We can also induce a current by keeping the secondary still and moving the primary. This is a variant of category (b). But now, we can induce current in the secondary in yet another way, namely by having a variable current in the primary. We can do that by opening or closing the switch or, for example, by including a variable resistor in the primary and changing its resistance quickly. Analyzing the case, we see that, in this way, the induction is also due to being the induced circuit in a magnetic field varying with time.

We can then conclude that the three above categories are really only two. In both cases, there is a circuit in a magnetic field. The electromagnetic induction phenomenon happens when the field is steady and the circuit moves or when the circuit is at rest and the magnetic field varies with time (what matters is the field variation; whether this is due to the motion of its sources or to a variation of the current that produces it is irrelevant, provided the variation is the same). We observe here that

the directly induced quantity is not the current, but rather an electromotive force (emf) along the circuit. As the reader remembers, the emf is the integral about the circuit (or a segment of the circuit) of the tangent component of the force on the unit charge. Let Γ be the oriented geometrical curve representing the circuit and \mathbf{F} the force on the charge q (which is, in general, a function both of the point and of time). The emf is, by definition,

$$\mathcal{E} = \int_{\Gamma} \frac{\mathbf{F}}{q} \cdot d\mathbf{s}. \quad (7.1)$$

To observe what we just stated, we must change the experimental setup a bit. We remove the galvanometer from the circuit and observe the phenomenon with an *open circuit*. In addition, we take the two ends of the circuit in a region in which there is no field varying in time. Under such conditions, the electrostatic potential difference is defined and we measure it with a voltmeter. Repeating the above-mentioned experiments with this setup, we observe that the observed potential difference is equal to the emf induced in the circuit. When the circuit was closed, the current intensity was given, for Ohm's law, by the ratio of the emf and the resistance of the circuit. Faraday established that, in all the cases, the following *flux rule* holds. The rule states that the emf induced about the circuit Γ is equal to the opposite of the rate of change of the magnetic flux linked to Γ ,

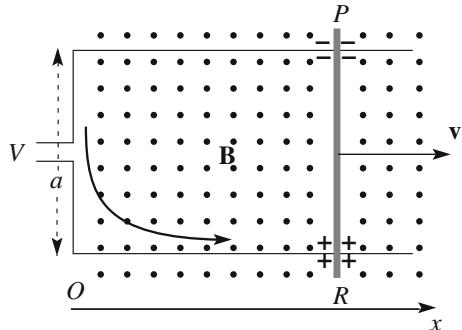
$$\mathcal{E} = -\frac{d\Phi}{dt}, \quad (7.2)$$

where Φ is the flux of the magnetic field through any surface Σ bounded by Γ namely

$$\Phi = \int_{\Sigma} \mathbf{B} \cdot \mathbf{n} d\Sigma. \quad (7.3)$$

Note that, in order to define the emf completely, we need to (arbitrarily) choose a positive direction on Γ . Similarly, the definition of the flux requires us to have chosen a positive face of Σ , namely the positive direction of the unit normal vector \mathbf{n} . The second choice is not arbitrary. Rather, the positive direction of \mathbf{n} is such that one observer with his feet on Σ having that direction entering through his feet and flowing out of his head sees the positive direction of Γ as counter-clockwise. We further notice that the flux of \mathbf{B} is the same through any oriented surface Σ bounded by Γ . This is a consequence of \mathbf{B} being a solenoidal vector, namely having identically null divergence. This demonstration is identical to that which we made in Sect. 6.6 for the flux of the current density \mathbf{j} under steady conditions, in which \mathbf{j} is solenoidal. The flux Φ , hence, depends only on Γ and is said to be *linked* to Γ .

Fig. 7.1 Circuit with a segment in motion in uniform and a steady magnetic field, directed perpendicularly to the figure above. The arrow marks the positive orientation of the circuit (not necessarily the direction of the current flow)



It is important to understand that the direct effect of the magnetic flux variations is the induced emf. The electric current is, in turn, an effect of the induced emf, provided the circuit is closed. If, contrastingly, the circuit is open, the emf exists, but there is no current. Concerning the circuit, the emf depends only on its geometry, while the current intensity also depends on its physical characteristics, in particular, its resistance R . Indeed, we have

$$I = \frac{\mathcal{E}}{R} = -\frac{1}{R} \frac{d\Phi}{dt}. \quad (7.4)$$

Another important observation is that the induced emf is not localized at a point of the circuit, as is the case with a battery, but is distributed along the entire length of the circuit, or at least of one segment, in a case in which this is the only part moving, such as the one we shall now consider in detail (Fig. 7.1).

7.2 Induced Electric Field. Faraday's Law

We shall now see how the two fundamental cases of the flux rule have completely different theoretical interpretations. In case 1, namely when the flux varies due to the motion of the parts of the circuit in the magnetic field, the rule is a consequence of the Lorentz force, namely of physics we already know. Contrastingly, case 2 shall require new physics, namely the modification of a law that we know to hold under static conditions. Let us start with case 1.

Let us consider the simple circuit in Fig. 7.1. It is made of two parallel metallic rails separated by a distance a . On the left side, the rails are almost, but not completely, joined by two conductors perpendicularly. We measure the potential difference between the extremes of these two conductors with a voltmeter, at open circuit. A metal bar lies on the rails and moves, remaining perpendicular to the rails with constant velocity v . We take the x -axis parallel to the rails in the direction of v with origin at the left end point of the rail. A uniform and constant magnetic field \mathbf{B} is present between the rails, but not at the voltmeter. Its direction is perpendicular

to the plane of the rails, outward in the drawing. We choose the positive direction along the circuit to have the linked flux of \mathbf{B} positive, as shown in Fig. 7.1.

Consider a generic charge carrier in the moving bar (an electron, in this case). Let q be its charge. The Lorentz force acts upon the carrier, which moves with the bar at velocity v :

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B}. \quad (7.5)$$

\mathbf{v} and \mathbf{B} being orthogonal to one another, and both perpendicular to the bar, the force has magnitude qvB and the direction of the bar. Note that the force is present at all the points of the moving bar and at no other point of the circuit. The force *per unit charge* of magnitude vB is directed toward the extreme P in the figure, independently of the charge of the carrier. For Eq. (7.1), the induced emf is then

$$\mathcal{E} = -vBa.$$

The physical process is such that the free electrons are pushed by the Lorentz force toward the extreme P in the figure, originating an accumulation of negative charge near P and one of positive charge near the other extreme R . These two charge accumulations generate an electric field \mathbf{E} directed along the bar from R to P , which repels further arrivals. The equilibrium is reached when the electric and magnetic forces on the carriers are equal and opposite at all the points of the bar, namely for

$$E = -vB.$$

The electrostatic potential difference between the extremes of the bars is then

$$\Delta V = Ea = -vBa, \quad (7.6)$$

which is equal to the induced emf.

Let us now calculate the linked flux. When the bar is at x , the flux is $\Phi = Bax$. Hence, its rate of change, taking into account Eq. (7.6), is

$$\frac{d\Phi}{dt} = Ba \frac{dx}{dt} = Bav = -\mathcal{E}$$

We have obtained the flux rule.

As we have seen, the direct effect is the emf. When we work with a closed circuit, this emf, in turn, originates a current, whose intensity depends on the resistance of the circuit R , according to Ohm's law, namely

$$I = -\frac{vBa}{R}. \quad (7.7)$$

Notice the minus sign. It indicates that the current circulates in a direction opposite to that which we have chosen as positive. The current is due to the emf at the extremes of the moving bar. Consequently, its direction along the fixed part of the circuit is from the positive end of the bar, which is R , to the negative end, namely P . The current direction is then counter-clockwise. Notice that the direction of the current in the moving bar is from the negative end to the positive end. This behavior is analogous to that of an electrostatic generator or a common battery. In this case, the non-electrostatic force inside the emf generator, namely the moving bar, is clearly the Lorentz force.

The induced current generates its own magnetic field. This field is directed perpendicularly to the page of the drawing toward the inside, namely parallel and opposite to the applied field. The field of the induced current contributes in this way to the total magnetic flux linked to the circuit, with a negative contribution, namely diminishing the flux of the applied field. The net effect is that of diminishing the flux grow rate due to the bar motion. If we now consider the bar moving to the left, namely reducing the linked flux, we find that the magnetic field produced by the induced current has the same positive direction as the applied field. The contribution of the induced current is then to increase the linked flux, or, we can say, to oppose the diminishment of the total flux. These observations hold, in general, every time the flux law holds, both in case 1 and in case 2. This property is known as the Lenz law, from Heinrich Lenz (Russian empire, 1804–1865), who formulated it in 1834, stating that *the direction of the induced current is always such so as to oppose the cause from which it has been generated*.

In conclusion, we can state that, in the case we considered, the flux rule is a consequence of the Lorentz force. It can be shown that the conclusion is valid whenever the emf is induced in a moving circuit.

Let us now analyze case 2. Consider a circuit, Γ , at rest. The charge carriers are at rest as well. The flux of the magnetic field \mathbf{B} linked to the circuit varies with time (The vector \mathbf{B} is a function of the coordinates and time.). No magnetic force acts on the carriers, because they are at rest. The observed effect must then be due to an electric field \mathbf{E} , producing action on the carriers. We are induced to think that the variations in time of the magnetic field give origin to an electric field and, consequently, to an emf, such as

$$\mathcal{E} = \oint_{\Gamma} \frac{\mathbf{F}}{q} \cdot d\mathbf{s} = \oint_{\Gamma} \mathbf{E} \cdot d\mathbf{s} = -\frac{d\Phi}{dt}. \quad (7.8)$$

We immediately see that the electric field is not conservative under these conditions. Indeed, its line integral about a closed line (namely the work done by the field on the unit charge moving on that loop) is not zero, but

$$\oint_{\Gamma} \mathbf{E} \cdot d\mathbf{s} = -\frac{d}{dt} \int_{\Sigma} \mathbf{B} \cdot \mathbf{n} d\Sigma, \quad (7.9)$$

where Σ is an arbitrary surface bounded by Γ . The same conclusion can be expressed in differential form, applying the Stokes theorem to the left-hand side of Eq. (7.9) to transform it into a surface integral over Σ , namely

$$\oint_{\Gamma} \mathbf{E} \cdot d\mathbf{s} = -\frac{d}{dt} \int_{\Sigma} \nabla \times \mathbf{E} \cdot \mathbf{n} d\Sigma = -\int_{\Sigma} \frac{\partial \mathbf{B}}{\partial t} \cdot \mathbf{n} d\Sigma,$$

where, on the right-hand side, we could bring the time derivative under the integral, because the integration limits (namely Σ) do not depend on time.

Being that the equation is valid for any Σ , the equality must hold for the integrands, and we have

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad (7.10)$$

which is a fundamental equation that we call the *Faraday law*. The law is universally valid and expresses the properties of the electric field together with the equation for the divergence:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}, \quad (7.11)$$

which is also universally valid.

The Faraday law extends the equation $\nabla \times \mathbf{E} = 0$ valid in electrostatics to dynamics, and to that it reduces when \mathbf{B} is constant in time. We see that there are two causes of an electric field, namely the electric charges, which determine its divergence, and the time variations of the magnetic field, which determine its curl. Equations (7.10) and (7.11) are *local* expressions. Namely, to know the divergence and curl of the electric field at a point in a certain instant, we just need to know the charge density and the rate of change of the magnetic field at that point and in that instant.

As we have noted, the electric field under non-static conditions is not conservative. Consequently, we cannot define a potential as we did with electrostatics. The emf takes the place of the potential difference under dynamic situations. However, as opposed to the potential difference, the emf is not a function of state. Indeed, the work done by the forces of the field on the unit charge to go from state A to state B does not depend on A and B alone, but also on the path followed. This fact, however, does not imply that the electromagnetic forces are dissipative. As a matter of fact, we shall subsequently see that not only matter but also the fields contain energy. Neither the matter nor the field energy is separately conserved, but their sum is, in any case, constant.

We now make an important observation on the Faraday law in integral form (Eq. 7.9). This equation holds for any oriented curve Γ , even if it is purely geometric and not materialized in a physical circuit. An electric field exists at the points of the curve, which is such that the integral about Γ of its tangent component is

equal and opposite to the rate of change of the flux of the magnetic field linked to Γ . We shall see an example of that in Sect. 7.4.

Let us go back, in conclusion, to the flux rule. As we have seen, the induced emf in the two cases, the case of the moving circuit and the case of the variable magnetic field, is due to two completely different causes. One is in the magnetic and one in the electric term in the force

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \quad (7.12)$$

The flux law is a *single* physics law resulting from *two different* physical phenomena. In similar cases, as purposefully Feynman stated, “such a beautiful generalization is found to stem from a single deep underlying principle. Nevertheless, in this case there does not appear to be any such profound implication”.

7.3 Exceptions to the Flux Rule

In the previous sections, we started from the experimentally determined *flux rule* to infer the Faraday law (Eq. 7.10). However, the universally valid laws are Eqs. (7.10) and (7.11) for the fields and Eq. (7.12) for the force. The flux rule is a consequence of them. Let us make the point clear by considering a few examples in which the flux rule cannot be applied. These exceptions happen when one deals with an extended conductor (namely not a wire) or when the path of the current in a conductor varies with time.

Barlow's wheel is a demonstration device invented in 1822 by Peter Barlow (UK, 1776–1862), shown in Fig. 7.2. The wheel is a vertical metal disk turning about a horizontal metal hub. Two wires are connected to the moving part with sliding contacts. One contact touches the hub, while the other is a mercury contact at the lowest point of the rim, which touches the surface of liquid mercury in a container not shown in the figure. The emf between the two wires is measured by a voltmeter. When we rotate the wheel with a certain angular velocity ω , we observe the presence of an emf. However, there is no magnetic flux variation. The explanation is as follows. The free electrons in the disk travel along the radius from the lowest point to the center with a speed that, at the distance r from the axis, is $v = \omega r$.

Fig. 7.2 Barlow's wheel

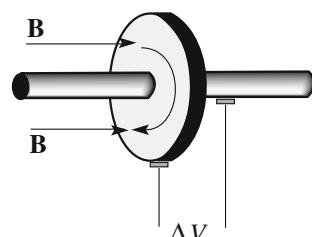
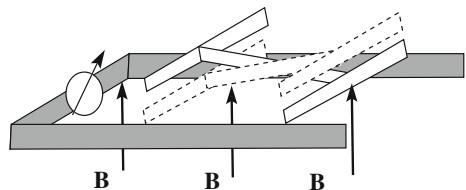


Fig. 7.3 Demonstration in which the flux varies without inducing an emf



As a consequence, they are acted upon by a magnetic force of intensity $F = qvB = q\omega rB$ (considering that v is perpendicular to \mathbf{B}) in the direction of the radius. The emf between the hub and the rim is

$$\mathcal{E} = \int_0^P \omega r B dr = \frac{1}{2} \omega B P^2, \quad (7.13)$$

where P is the wheel's radius. We see that the flux rule does not hold, because the circuit, if meant to be the geometric path of the current, does not vary, and consequently there is no change in the linked flux. Contrastingly, when considered as the physical support to the current, the circuit does vary, because the current-carrying instantaneous radius varies, and consequently varies the part on which the magnetic force is acting.

Figure 7.3 shows an opposite example. The device in this case is made of two conducting parallel rails joined by a third one. We also have a conductor made by joining three conducting bars in the shape of an H. A constant and uniform magnetic field is present normal to the plane of the rails. If we touch the two rails with an arm of the H, we close the circuit. We then raise that arm and join the rails with the second arm, as in the figure. In the process, the area of the circuit varies considerably, and so does the linked magnetic flux, but no induced emf is observed.

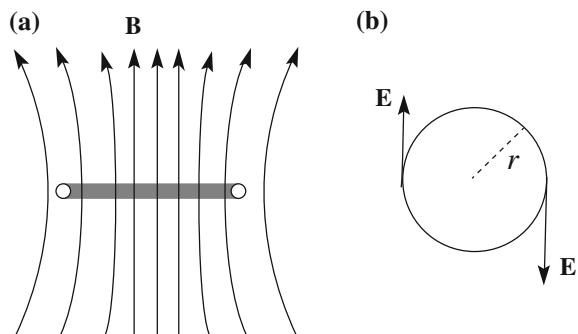
The reason for this is that, when we move the H, the free electrons inside it move, but their velocities are substantially parallel to the magnetic field, and consequently the Lorentz force is null.

In conclusion, we must pay attention when using the flux rule under any circumstance in which the matter through which the current flows changes. We can be sure, on the other hand, both of the Faraday law (Eq. 7.10) and of the expression of the force (Eq. 7.12).

7.4 Betatron

We shall now discuss an example of electromagnetic induction in which a material circuit does not exist at all, the betatron. As we already stated, Eq. (7.9) also holds if Γ is a purely geometric curve. Donald Kerst (USA, 1911–1993) built the first magnetic induction particle accelerator at the University of Illinois in 1940, which

Fig. 7.4 Scheme of a betatron. **a** Vertical cross-section on axis, **b** the electric field line, which is the trajectory of a particle



was called a betatron. Betatrons accelerate charged particles, typically electrons (hence the name), up to quite high energies. We are mainly interested here in the working principles, and we shall consequently ignore all technical details.

The main component of the betatron is an electromagnet that generates a magnetic field with cylindrical symmetry, as shown in Fig. 7.4a. The field varies in time, as needed. A pipe having the shape of a donut is located coaxially in the middle plane, showed cut in Fig. 7.4a. The donut is under vacuum to let the electrons circulate freely. During the acceleration process, the magnetic field intensity increases. The increasing flux linked to the donut induces the accelerating electric field. The electric field lines are circles centered on the axis, as symmetry requires.

The orbit of the electrons is one of these circles (we shall soon see which of them it is). Let it be Γ and let r be its radius. The flux linked with Γ can be written as the mean value of the magnetic field inside the circle of radius r , say $\langle B(r, t) \rangle$, times the area of the circle. Namely, the flux is $\Phi = \pi r^2 \langle B(r, t) \rangle$. Equation (7.9) then gives us

$$2\pi r E(r, t) = -\pi r^2 \frac{d\langle B(r, t) \rangle}{dt},$$

or, in absolute values,

$$|E(r, t)| = \frac{r}{2} \left| \frac{d\langle B(r, t) \rangle}{dt} \right|. \quad (7.14)$$

The force acting on one electron is $q_e E$, where q_e is the electron charge, and is proportional to the rate of change of the (mean) magnetic field. The force is tangent to the orbit, and consequently makes the magnitude of the momentum p vary at the rate of

$$\frac{dp}{dt} = q_e E = \frac{q_e r}{2} \left| \frac{d\langle B \rangle}{dt} \right|. \quad (7.15)$$

If the electron is initially ($t = 0$) at rest and the field is initially zero, the momentum of the electron at the generic time t is

$$p(t) = \frac{q_e r}{2} \int_0^t \frac{d\langle B \rangle}{dt} dt = \frac{q_e r}{2} \langle B(r, t) \rangle. \quad (7.16)$$

In this way, we can increase the momentum, and consequently the kinetic energy, of the electrons by increasing the magnetic field. We still need to see how we can keep the electrons moving in a circular orbit. Indeed, for that, we need a centripetal force with the right value in every moment. As a matter of fact, a centripetal force does exist. It is the Lorentz force $q_e \mathbf{v} \times \mathbf{B}$. We know from mechanics that the centripetal force is equal to the product of the p and the angular velocity $\omega = v/r$. Namely, we have the condition $qvB(r, t) = pv/r$, which is

$$B(r, t) = \frac{1}{2} \langle B(r, t) \rangle. \quad (7.17)$$

The condition means that, at any time, the field on the orbit must be equal to one half of its mean value inside the same orbit. This condition is obtained by properly shaping the poles of the electromagnet.

Notice that the mass of the electron does not appear in any of the expressions we have used. As a matter of fact, the betatron can accelerate electrons up to hundreds of MeV, namely at highly relativistic energies, much larger than the electron rest energy, which is about half a MeV. The upper limit comes from the fact that accelerated charged particles (centripetal acceleration, in this case) radiate electromagnetic energy. This phenomenon is called synchrotron radiation, because it happens in synchrotrons as well. The radiated power increases as the fourth power of the energy of the particles. While synchrotron radiation is an undesired effect for an accelerating particle, it too may be very useful. Being that the radiation is mainly in the X-ray range, betatrons of about 10 MeV energy are designed as “hard” X-ray sources for industrial and medical radiography.

We shall now express the same concepts in terms of the vector potential. Note that, even if we have defined the vector potential in magnetostatics, the definition holds under dynamic conditions as well. This is because it is based on $\nabla \cdot \mathbf{B} = 0$, which continues to hold in dynamics, as we shall see in Chap. 10.

Let us express Eq. (7.9) on the trajectory Γ of the electrons. We notice that this circle is a line of the vector potential as well. On the other hand, the flux of \mathbf{B} through the circle bounded by Γ is equal to the circulation of \mathbf{A} about Γ (see Eq. 6.46) and this is $2\pi r A(r, t)$. We then have

$$2\pi r E(r, t) = -\frac{d\Phi_B}{dt} = -2\pi r \frac{\partial A(r, t)}{\partial t},$$

namely also

$$E(r, t) = -\frac{\partial A(r, t)}{\partial t}.$$

The rate of change of the electron momentum is

$$\frac{dp}{dt} = q_e E(r, t) = -q_e \frac{\partial A(r, t)}{\partial t}.$$

We now integrate over time, taking into account that at $t = 0$, both p and A are zero, obtaining

$$p(t) = -q_e A(r, t), \quad (7.18)$$

which can be also written as

$$p(t) + q_e A(r, t) = \text{constant}. \quad (7.19)$$

This simple relation holds in every instant. It tells us that the sum of the momentum of the electron and the product of its charge times the vector potential is a constant of the motion. We shall come back to this issue in Chap. 10.

7.5 Felici's Law

Felici's law allows for calculating the net charge through a circuit due to a current induced by a variable magnetic field knowing just the initial and final magnetic flux and the resistance of the circuit. The law was discovered by Riccardo Felici (Italy 1819–1902) in his systematic experimental work on electromagnetic induction carried on between 1851 and 1859. Today, we can obtain the law as a consequence of the flux law by integrating Eq. (7.4) over a given time interval, say from t_1 to t_2 . Taking onto account that the integral of the current intensity over a time interval is the net charge through the conductor in that interval, which we call Q , we immediately have

$$Q = \frac{\Phi_1 - \Phi_2}{R}, \quad (7.20)$$

where R is the resistance of the circuit and Φ_1 and Φ_2 are the magnetic fluxes linked to the circuit at t_1 and t_2 , respectively. This is Felici's law.

The interesting point is that the net charge does not depend on the specific evolution in time of the magnetic flux, but only on its total change. This observation suggests a way to measure the flux, and consequently the magnetic field. It is called the flip-coil method. We wind a copper wire producing a coil of n overlapped loops.

We make its area, which we call S , small enough to be able to consider the magnetic field to be measured as uniform on that area. We connect the coil to an instrument suitable for measuring the total electric charge, namely to integrate the current. An example of such a device is the ballistic galvanometer, which we shall describe.

Suppose we know the direction of the magnetic field. (We shall later address what to do if we do not.) We insert the coil at the point of interest with its surface perpendicular to the field. The initial flux is $\Phi_1 = nSB$. We now quickly flip the coil by 180° , changing the flux to $\Phi_2 = -nSB$. For Felici's law, the magnetic field is

$$B = \frac{QR}{2nS}. \quad (7.21)$$

If we do not know the direction of the field, we can proceed by trial and error, repeating the procedure but starting with the coil in different directions and finding the one for which the induced charge Q is at a maximum.

Let us now see how to measure the total charge with a ballistic galvanometer. This instrument is just a galvanometer, as described in Sect. 6.5, whose mobile coil inertia moment, J , is quite large. Consequently, its oscillation period is long compared to the total time in which the current flows. We recall that the coil carrying the current I is subject to a torque

$$\tau_I = IB_c S_c N_c. \quad (7.22)$$

where B_c is the magnetic field on the coil, S_c the area of the coil and N_c the number of its loops. This torque produces an angular acceleration of the coil given by

$$\tau_I = J \frac{d\omega}{dt}, \quad (7.23)$$

where ω is its angular velocity. In the generic infinitesimal time interval dt , the angular impulse is

$$Jd\omega = \tau_I dt = IB_c S_c N_c dt. \quad (7.24)$$

We can talk of an angular pulse because the duration of the current is small compared to the period. Integrating from the initial instant ($t = 0$) to the final one t_c , when the entire current has flowed, we have

$$\int_0^{\omega_0} J d\omega = B_c S_c N_c \int_0^{t_c} I dt = B_c S_c N_c Q. \quad (7.25)$$

This means that the final angular velocity of the coil, ω_0 , is proportional to the total charge carried by the current, with a proportionality constant K that we do not need to express

$$\omega_0 = KQ. \quad (7.26)$$

After the current has ceased, namely for $t > t_c$, the galvanometer coil freely oscillates under the action of its spring, only with initial angular velocity ω_0 . The mechanical energy of the coil is initially purely kinetic and equal to $J\omega_0^2/2$. The coil then turns, reaching a maximum angle, say α_m , where it momentarily stops. Here, all the initial kinetic energy becomes potential energy. The latter being proportional to α_m^2 , we can conclude that the maximum coil elongation α_m is proportional to ω_0 , and hence to Q , namely that

$$\alpha_m = K'Q. \quad (7.27)$$

Measuring α_m , we shall know Q , once we know the constant K' . We might easily express it in terms of the characteristics of the coil and the spring. In practice, this method does not provide a sufficient precision, and one prefers to calibrate the galvanometer by measuring a number of maximum elongations α_m for charges of known values.

The ballistic galvanometer is similar to the ballistic pendulum discussed in Sect. 7.18 of the 1st volume of this course.

QUESTION Q 7.1. A coil of area $A = 1 \text{ cm}^2$ and resistance $R = 100 \text{ m}\Omega$ is located in a field $B = 1.2 \text{ T}$ directed perpendicularly to it. We quickly turn the coil upside down (by 180°). What is the charge crossing the coil? \square

7.6 Energy Balance

Let us go back to the example of the mobile side circuit we considered in Sect. 7.2. Suppose the friction of the motion of the mobile bar to be negligible. Suppose the circuit to be open and the bar initially to be moving at a speed v . Under these conditions, a magnetic force is present on the bar. Being parallel to the bar, it generates an emf but does not have an effect on the velocity, to which it is perpendicular. If we now close the circuit on a resistance R , the current of intensity given by Eq. (7.7) flows into the circuit, mobile segment included. Each element $d\mathbf{s}$ of the mobile side is acted upon by the force

$$d\mathbf{F} = Id\mathbf{s} \times \mathbf{B}.$$

This force is parallel and opposed to the bar velocity (remember Lenz's law). The total force on the bar is

$$F = IBa = -\frac{B^2 a^2}{R}v, \quad (7.28)$$

where a is the length of the bar and the minus sign indicates that the force is opposite to v . We see that, when the circuit is closed, the force acting on the mobile side has characteristics similar to the viscous drag, namely it is proportional and opposite to the velocity. The force slows down the bar until it stops. Where has the initial mechanical energy of the bar gone? It has been dissipated by the Joule effect on the resistance R of the circuit. Indeed, the dissipated electric power is

$$w_{\text{el}} = RI^2 = B^2 a^2 v^2 / R. \quad (7.29)$$

If we want the bar moving at a constant speed, we must apply a force equal and opposite to F . The force provides a mechanical power equal to

$$w_{\text{mec}} = Fv = B^2 a^2 v^2 / R, \quad (7.30)$$

which, as we see, compensates the electrical power. Such a device is an elementary prototype of an electromagnetic generator, in which mechanical work is spent to produce an electric current when the circuit is closed on a *load* R (it might be the resistance of a lamp, for example). The mechanic work spent is equal to the electric work obtained, when dissipative losses can be neglected.

Similar considerations hold for Barlow's wheel, as described in Sect. 7.3. In that case as well, when the circuit is open, there are no mechanical moments on the wheel, if we can neglect friction. If initially rotating, the wheel will continue to turn at a constant angular velocity. When the circuit is closed on a total resistance R , a current flows on the radius of the wheel between the hub to the mercury contact. Its intensity is

$$I = \frac{\mathcal{E}}{R} = \frac{\omega B P^2}{4R}, \quad (7.31)$$

where P is the radius of the wheel. Under these conditions, a force $dF = IBdr$ acts on the generic element dr of the radius, directed perpendicularly to the radius. Its moment about the axis is $d\tau = IBrdr$. We obtain the total moment by integration on the radius, namely

$$\tau = \frac{IBP^2}{2} = -\frac{B^2 P^4}{2R} \omega, \quad (7.32)$$

where the minus sign indicates that its direction is opposed to the angular velocity (again, Lenz's law). In this case too, the acting moment is similar to a viscous moment. Consequently, there is a loss of mechanical energy per unit time

$$w_{\text{mec}} = \tau \omega = \frac{B^2 P^4 \omega^2}{4R}. \quad (7.33)$$

Once more, the mechanical power is equal to the electrical one. Indeed, it is

$$w_{\text{el}} = RI^2 = \frac{B^2 P^4 \omega^2}{4R}. \quad (7.34)$$

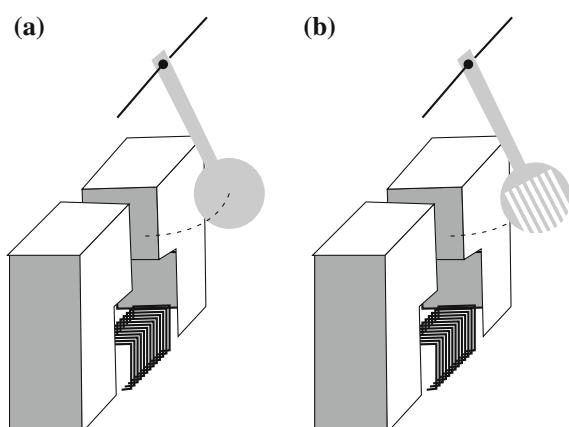
If we want to keep the angular velocity constant, we must provide a mechanical moment equal and opposite to that of Eq. (7.32), and consequently spend mechanical power equal to the electrical power (or larger in the presence of friction).

Barlow's wheel can be considered to be another example (even without practical interest) of an electromagnetic generator, namely a device transforming mechanical work into electric work. It is interesting to note that the device can also be used as an electric motor, namely to transform electric work into mechanical work. To obtain that, we just have to include a source of emf, which will provide the electrical work, in the circuit. When the circuit is closed, the emf generates a current, whose intensity we call I . Consequently, a force $d\mathbf{F} = -Id\mathbf{r} \times \mathbf{B}$ acts on every segment $d\mathbf{r}$ of the radius. Its moment about the axis is $d\tau = \mathbf{r} \times d\mathbf{F} = -Ir \times (\mathbf{B} \times d\mathbf{r})$. Taking into account that \mathbf{B} is perpendicular to \mathbf{r} , its magnitude is $d\tau = IBRdr$. The total mechanical moment about the axis is obtained by integration from 0 to P , the radius of the wheel, obtaining $\tau = IBP^2/2$. We can use this moment for mechanical work.

7.7 Eddy Currents

Consider the experimental demonstration shown in Fig. 7.5. A pendulum is made of a small metal bar joined to a metal disk, which moves between the poles of an electromagnet around the equilibrium position. When the magnet is off, the pendulum oscillates as usual. When we excite the magnet, namely produce a strong

Fig. 7.5 Metal pendulum swimming between the poles of a magnet. Eddy currents are strong in (a), small in (b)



magnetic field between its poles, the damping of the oscillation increases substantially, as if the viscous drag would have enormously increased.

The greater the speed of the pendulum is, the more striking the effect. To understand this phenomenon, think of the moments in which the pendulum disk penetrates the field region from outside. As it goes in, the flux linked to possible circuits on its surface increases. The linked flux increase generates electromotive forces that, in turn, produce eddies of current on the disk. The presence of these eddy currents, as they are called, explains, similarly to what we discussed in the last section, the drag resistance proportional and opposite to the velocity. The eddy currents and the corresponding forces may be quite large due to the very low electric resistance of the “circuits” on the metal disk. If the disk were a perfect conductor, the currents would be so intense that they would push the disk back out.

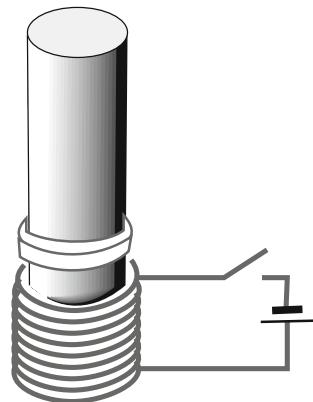
We can check that this interpretation is correct by repeating the experiment with a pendulum equal to the first one, but having a series of cuts perpendicular to the direction of motion, as in Fig. 7.5b. We observe that the dragging force practically disappears. Independently of the field being on or not, the behavior of the oscillations is similar. Indeed, the cuts forbid the formation of current eddies.

The case discussed is just an example of a rather common phenomenon. In the presence of magnetic flux variations, eddy currents, also called Foucault currents, always develop. Their consequence is energy dissipation. Consequently, measures must be taken to reduce the losses to a minimum. Consider, for example, the transformers used in electric technology to “transform” an alternate current from one voltage to another. In a transformer, two circuits made of a different number of loops are wound around the same cylindrical iron structure, called the core. The alternate current in one of the circuits, the primary, induces another alternate current in the secondary. Consequently, a magnetic field varying in time is present in the core. To minimize the energy losses, the core is made of laminated sheets packets, arranged in the direction perpendicular to the eddy currents.

Let us consider two other examples.

Figure 7.6 shows the Thomson jumping ring experiment conducted by Elihu Thomson (UK-USA, 1853–1937) in the 1880s. The device is made of a metal ring inserted into a vertical insulating bar on the axis of an electromagnet, for example, on the face of a solenoid. Initially, the magnet is off and the ring rests on its face. When we switch the magnet on, we see the ring jumping up (independently of the direction of the current). This phenomenon is due to the eddy currents that develop in the ring when the magnetic field increases. These currents generate a magnetic field, which, for Lenz’s law, opposes the causes that generated it. Hence, its direction is opposite to that of the original field, and the ring is pushed up. When the field has reached its steady value, the currents soon disappear due to the resistance of the ring, and the ring falls back into its original position. If we now switch the magnet off, we see the ring pressed onto the magnet’s face. It is trying to keep linked to as much magnetic flux as possible, to obey Lenz’s law. We can check our

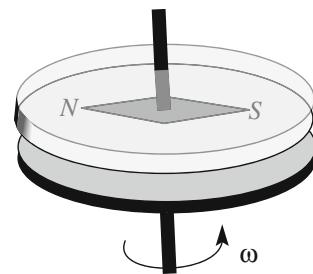
Fig. 7.6 The Thomson jumping ring



interpretation by repeating the experiment with a cut ring. We do not observe any effect.

Figure 7.7 shows an earlier experiment done by François Arago (France, 1786–1853) in 1826. This was probably the first historic observation of an electromagnetic induction phenomenon, but was correctly interpreted only after Faraday's discoveries. In the experiment, a magnetic needle is enclosed in a transparent box, closed tightly to avoid any effect linked to movements of air. The needle is free to turn on a vertical hub. A coaxial copper disk, parallel to the box at a short distance below it, can be put in rapid rotation. When we do so, we observe the needle turning as well, as if the disk is dragging it, even if there is no mechanical interaction between them. This effect is due to the eddy currents. Initially, when the copper disk moves and the needle is still at rest, there is a relative motion between them. The magnetic field of the needle appears to an observer of the disk to be varying in time. This generates eddy currents in the copper. The action of the currents is to oppose the cause that generated them, which is ultimately the relative motion. Consequently, the needle rotates.

Fig. 7.7 The Arago disk



7.8 Mutual Induction

Let Γ_1 and Γ_2 be two circuits at rest and let us orient them arbitrarily. Let I_1 be the current carried by the first circuit and \mathbf{B}_1 the magnetic field generated by that current. Let Φ_{12} be the flux of \mathbf{B}_1 linked to the second circuit, namely Γ_2 . Any variation in time of I_1 will generate an emf in Γ_2 . Clearly, we can also consider that a current I_2 in Γ_2 generates a magnetic field \mathbf{B}_2 and a flux Φ_{21} of \mathbf{B}_2 linked to Γ_1 . Any variation in time of I_2 will generate an emf in Γ_1 . This is the phenomenon of mutual induction, which we shall now study.

Figure 7.8 defines the geometrical quantities. In addition, let Σ_1 and Σ_2 be surfaces bounded by Γ_1 and Γ_2 , respectively oriented according to the orientation of the corresponding curve, and \mathbf{n} the positive unit vector. The flux Φ_{12} is given by

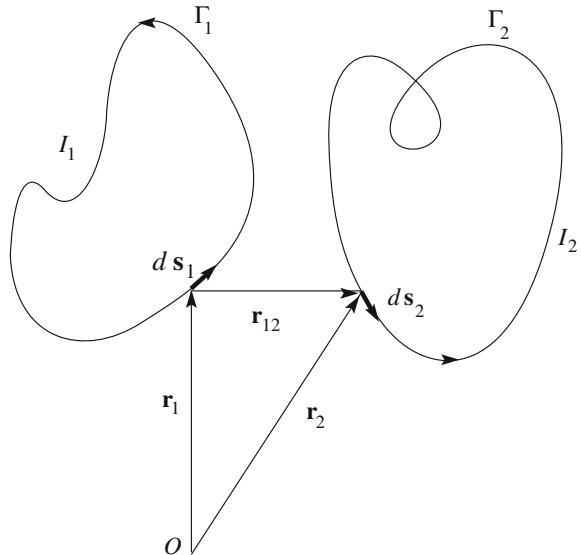
$$\Phi_{12} = \int_{\Sigma_2} \mathbf{B}_1(\mathbf{r}_2) \cdot \mathbf{n} d\Sigma. \quad (7.35)$$

Knowing that $\mathbf{B}_1 = \nabla \times \mathbf{A}_1$ and using the Stokes theorem, we have

$$\Phi_{12} = \int_{\Sigma_2} \mathbf{B}_1(\mathbf{r}_2) \cdot \mathbf{n} d\Sigma = \oint_{\Gamma_2} \mathbf{A}_1(\mathbf{r}_2) \cdot d\mathbf{s}_2. \quad (7.36)$$

Let us now assume both circuits to have such narrow cross-sections that their thickness can be neglected compared to all the other lengths of the problem. Under this assumption, we can write the vector potential of the first circuit as

Fig. 7.8 Mutual induction between two circuits



$$\mathbf{A}_1(\mathbf{r}_2) = I_1 \frac{\mu_0}{4\pi} \oint_{\Gamma_1} \frac{d\mathbf{s}_1}{r_{12}}. \quad (7.37)$$

We see that the vector potential, and consequently the magnetic field \mathbf{B}_1 and the linked flux Φ_{12} , are proportional to the current intensity I_1 producing them. Indeed, putting the two equations together, we find

$$\Phi_{12} = I_1 \frac{\mu_0}{4\pi} \oint_{\Gamma_1} \oint_{\Gamma_2} \frac{d\mathbf{s}_1 \cdot d\mathbf{s}_2}{r_{12}} = I_1 M_{12}, \quad (7.38)$$

where we have defined the quantity M_{12} as the proportionality constant between the current intensity in circuit 1 and the magnetic flux linked to 2, namely

$$M_{12} = \frac{\mu_0}{4\pi} \oint_{\Gamma_1} \oint_{\Gamma_2} \frac{d\mathbf{s}_1 \cdot d\mathbf{s}_2}{r_{12}}. \quad (7.39)$$

Let us now suppose that circuit 2 carries the current I_2 , producing a magnetic flux Φ_{21} linked to circuit 1, which, by the same arguments, is proportional to I_2 . We can write

$$\Phi_{21} = M_{21} I_2. \quad (7.40)$$

Calculating M_{21} with the same argument as before, we shall find an expression equal to Eq. (7.39) with indices 1 and 2 inverted. But that expression is symmetric under this inversion and we conclude that

$$M_{12} = M_{21} = M. \quad (7.41)$$

This is called the *mutual induction coefficient* or *mutual inductance* and is a property of the pair of circuits, independent of which one of them induces and which is induced. Consequently, we have indicated it simply with M without subscripts. The measurement unit of the mutual inductance is the henry (H), in honor of Joseph Henry (USA, 1798–1878) and in recognition of his outstanding contributions to the discovery of electromagnetic induction. By definition, the mutual induction between two circuits is equal to one henry (1 H) when a current of 1 A in one of them produces a linked flux of 1 Wb in the other one. We notice that the mutual induction coefficient may be both positive and negative, depending on the choice, which is arbitrary, of the positive orientations of the circuits.

While the mutual induction *coefficient* is defined in regard to steady currents, as we have just done, the mutual induction *phenomenon* appears when the current intensities vary in time. Suppose, for example, that the current I_1 is variable. It will produce a varying magnetic flux linked to the circuit 2 and consequently an emf about it, namely

$$\mathcal{E}_2 = -M \frac{dI_1}{dt}. \quad (7.42)$$

Similarly, a current I_2 variable in circuit 2 produces, in circuit 1, the emf

$$\mathcal{E}_1 = -M \frac{dI_2}{dt}. \quad (7.43)$$

7.9 Self-induction

A current-carrying circuit generates, in any case, a magnetic field, and consequently a magnetic flux linked with the circuit itself, which is sometimes called a self-flux. The self-flux Φ is proportional to the current intensity I , and we write

$$\Phi = LI, \quad (7.44)$$

where the proportionality constant L is called the *coefficient of self-induction* or *self-inductance*, or simply *inductance*.

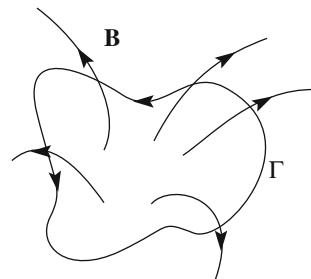
The measurement unit for the self-inductance in the SI is the henry, as it is for mutual inductance.

Let Γ be the oriented curve representing the circuit (the current I shall be positive if flowing in the positive direction, negative if in the opposite). Let Σ be any surface bounded by Γ oriented with the usual convention relative to Γ (See Fig. 7.9). The magnetic flux is

$$\Phi = \int_{\Sigma} \mathbf{B} \cdot \mathbf{n} d\Sigma. \quad (7.45)$$

We see that the self-inductance depends only on the geometry of the circuit. For a given current intensity, the self-flux, and hence the self-inductance, is that much larger the greater the size of the circuit it wraps. It is large for a solenoid with many

Fig. 7.9 A current-carrying circuit and the self-flux of its magnetic field



loops, small, for example, for a circuit consisting of two parallel wires close to one another. Note that the self-inductance is always positive, independent of the choice of a positive orientation of the circuit.

Note that an expression for the self-inductance similar to Eq. (7.39) for mutual inductance does not exist. Indeed, we found Eq. (7.39) assuming the points of both circuits always to be at distances large compared to the thickness of the circuits themselves. In this assumption we could use Eq. (7.37) to express the vector potential. Clearly, the assumption cannot be made in calculating the self-inductance. Rather, we must perform the integral in Eq. (7.45), as we shall do on a few examples.

The self-induction phenomenon appears when the current in the circuit varies with time. In this case, an emf appears about the circuit, which is

$$\mathcal{E} = -L \frac{dI}{dt}. \quad (7.46)$$

The emf, as we see, is proportional to the self-inductance and to the rate of change of the current. Notice that the induced emf contributes itself to the current in the circuit.

Let us calculate the self-inductance in a few geometrically simple examples.

A long solenoid. In Chap. 6, we calculated the magnetic field of a very long solenoid. Let N be the total number of loops, S the area of a loop, and l the length of the solenoid. The magnitude of the field is

$$B = \mu_0 N I / l. \quad (7.47)$$

The linked flux is N times the flux linked to a single loop, namely

$$\Phi = \mu_0 N^2 (S/l) I. \quad (7.48)$$

As expected, it is proportional to the current intensity, and we have

$$L = \mu_0 N^2 (S/l). \quad (7.49)$$

Let us look at the orders of magnitude. Consider, for example, a solenoid $l = 1$ m long made of $N = 10^4$ loops of 10 cm diameter. We can produce a magnetic field of 1 T inside it with a current $I = 80$ A. The inductance of the solenoid is $L = 0.99$ H, which is quite a large value, much larger than that of the circuit elements commonly used in electronic circuits.

Toroidal solenoid. In Chap. 6, we found that the magnetic field at the distance r from the axis of a torus of a square cross-section, radiiuses R_1 and R_2 and N loops is

$$B(r) = \mu_0 \frac{NI}{2\pi r}. \quad (7.50)$$

The total linked flux is N times the flux linked to a single loop. Let us calculate this flux, taking into account that the field is a function of r . The infinitesimal flux through an area element between r and $r + dr$ and having the height of the loop, namely $R_2 - R_1$, is

$$d\Phi = B(r)(R_2 - R_1)dr = \mu_0 \frac{NI}{2\pi} (R_2 - R_1) \frac{dr}{r}.$$

Integrating over r , we obtain the flux linked to one loop, namely

$$\Phi = \mu_0 \frac{NI}{2\pi} (R_2 - R_1) \ln \frac{R_2}{R_1}.$$

The self-inductance (remembering to multiply by N so as to have the total flux) is then

$$L = \mu_0 \frac{N^2}{2\pi} (R_2 - R_1) \ln \frac{R_2}{R_1}. \quad (7.51)$$

To get an idea of the orders of magnitude, consider a small toroidal solenoid (that we might use in a circuit) of $N = 500$ loops and radiiuses of 5 and 10 mm. Calculating the inductance, we find $L = 173 \mu\text{H}$.

Bifilar line. Circuits, or parts of circuits, made of two conducting and insulated wires running parallel to one another, carrying currents flowing in opposite directions, are quite common. Schematically, we can consider the two wires to be straight, indefinitely extended at a distance w . Let a be the diameter of the wires and I the current intensity, as shown in Fig. 7.10. Each of the wires produces a magnetic field at a distance r from its own axis equal to

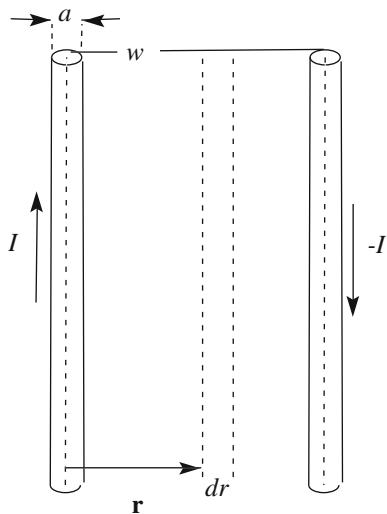
$$B = \mu_0 \frac{I}{2\pi r}. \quad (7.52)$$

The self-inductance of an infinitely long line is obviously infinite. Let us calculate the inductance per unit length L_u (measured in H/m).

Let us calculate the flux of the magnetic field produced by one of the wires, the one on the left, for example, as shown in Fig. 7.10, through the area between the wires of 1 m length. This will be the contribution of one wire. The contribution of the other being equal, we shall simply multiply it by two at the end. The flux of the field of the wire on the left-side in the area between r and $r + dr$ 1 m long is

$$d\Phi = B dr = \frac{\mu_0}{2\pi} I \frac{dr}{r}$$

Fig. 7.10 A section of bifilar line



We must integrate this between $a/2$ and $w-a/2$. We obtain

$$\Phi = \frac{\mu_0}{2\pi} I \int_{a/2}^{w-a/2} \frac{dr}{r} = \frac{\mu_0}{2\pi} I \ln \frac{w-a/2}{a/2}$$

Remembering to multiply by two, we finally have

$$L_u = \frac{\mu_0}{\pi} \ln \frac{w-a/2}{a/2}. \quad (7.53)$$

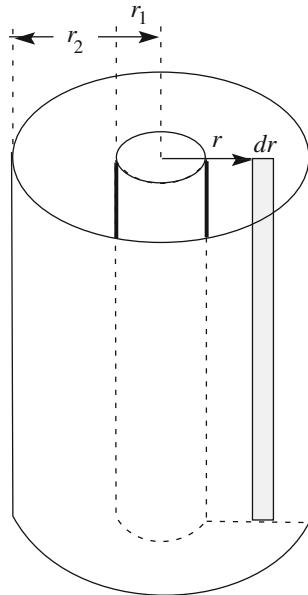
Let us look at the orders of magnitude. Consider wires of $a = 1$ mm diameter at a distance $w = 2$ mm. We find that the inductance per unit length is $L_u = 0.44 \mu\text{H/m}$.

QUESTION Q 7.2. A circuit consists of two parallel (insulated) wires carrying current in opposite directions. Consider two cases: (a) the wires are near and parallel, (b) the wires are twisted one about the other (as in a telephone line). In which case is the inductance larger? \square

Coaxial cable. A common configuration used, in practice, to transmit electric signals is the coaxial cable. It is made of two cylindrical coaxial metal conductors separated by an insulator, as shown in Fig. 7.11, carrying the same current in opposite directions. Let r_1 and r_2 be the radii of the two conductors.

The Ampère law tells us that the field in the region between the conductors is equal to the field of a straight current of the given intensity on the axis. For the same law, the field is zero outside the cable, because the two currents are equal and opposite. Let us express the self-inductance per unit length.

Fig. 7.11 A section of a coaxial cable



The flux through the area between r e $r + dr$ and 1 m long is

$$d\Phi = Bdr = \frac{\mu_0}{2\pi} I \frac{dr}{r}.$$

Integrating between r_1 and r_2 , we have

$$\Phi = \frac{\mu_0}{2\pi} I \int_{r_1}^{r_2} \frac{dr}{r} = \frac{\mu_0}{2\pi} I \ln \frac{r_2}{r_1},$$

which gives the self-inductance per unit length

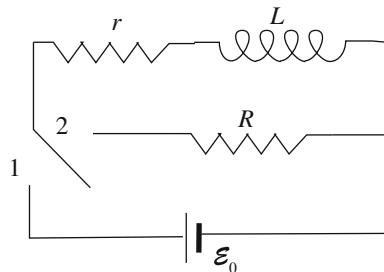
$$L_u = \frac{\mu_0}{2\pi} \ln \frac{r_2}{r_1}. \quad (7.54)$$

For example, the inductance of a 1 m long coaxial cable of radii 2 and 2.5 mm is $L_u = 0.45 \mu\text{H}/\text{m}$.

7.10 Inductive Phenomena in Electric Circuits

As we have seen, every time the current intensity varies in a circuit, an electro-motive force appears in it, given by

Fig. 7.12 A simple circuit with a constant emf generator



$$\mathcal{E} = -L \frac{dI}{dt}. \quad (7.55)$$

We recall that the minus sign on the right-hand side expresses Lenz's law. It tells us that when the current intensity diminishes, the induced emf is such that the current it produces flows in the same direction as the original current. In this way, it opposes the cause that has generated it. Similarly, if the current intensity increases, the current produced by the induced emf flows in the direction opposed to the original one. For this reason, the effect is called a *counter-electromotive force* (cemf, for short) or *back electromotive force*.

Induced emfs are always present in alternate current circuits, as we shall subsequently see, and also in circuits with constant emf generators alone, when the circuit is closed or open.

Let us consider the circuit shown in Fig. 7.12. It includes a constant emf generator (a battery, for example) \mathcal{E}_0 having negligible internal resistance, a resistor R , an inductor L (for example, a solenoid we want to use to produce a magnetic field) that has its own resistor r (drawn as a separate element in the figure, but physically a property of the inductance itself) and a switch. We first put the switch into position 1 to excite the magnetic field. The current intensity grows and a cemf develops according to Eq. (7.55). Ohm's law gives us

$$\mathcal{E}_0 - L \frac{dI}{dt} = rI. \quad (7.56)$$

This is a homogeneous differential equation with constant coefficients in the unknown function $I(t)$. We write the equation in the standard form

$$\frac{dI}{dt} + \frac{r}{L} I = \frac{\mathcal{E}_0}{L}. \quad (7.57)$$

Its general solution is the sum of a particular solution and the general solution of the associated homogenous differential equation. It is easy to find the former, considering that, in the steady state, namely after the initial transient is finished, the current intensity is constant and the cemf is zero. In the steady regime, we have

$$I = \frac{\mathcal{E}_0}{r}. \quad (7.58)$$

We find the solution to the associated homogeneous equation by separation of variables, writing it in the form of

$$\frac{dI}{I} = -\frac{r}{L} dt.$$

Integrating, we get

$$I(t) = Ae^{-(r/L)t},$$

where A is the integration constant. The solution to Eq. (7.57) is then

$$I(t) = Ae^{-(r/L)t} + \mathcal{E}_0/r.$$

We now determine the constant A by imposing the current as being zero at $t = 0$, and have the complete solution

$$I(t) = \frac{\mathcal{E}_0}{r} \left(1 - e^{-(r/L)t} \right). \quad (7.59)$$

We see that, starting from 0, the current intensity exponentially grows toward the value of the steady regime. Rigorously speaking, the time to reach the steady state would be infinite. In practice, it is enough to wait a few times for the constant in the exponent, namely $\tau = L/r$, which is called the *time constant* of the circuit. For example, if the circuit includes an electromagnet, L is quite large and r quite small, corresponding to a several-seconds-long time constant. The time constant is much shorter in a common electronic circuit.

Let us now shut down the current in the inductor, by moving the switch into position 2. The current intensity decreases and a cemf appears. Once more, Ohm's law gives us

$$-L \frac{dI}{dt} = (r + R)I.$$

The equation is now homogeneous. Let us take $t = 0$ in the moment we move the switch to position 2. The current intensity then has the steady state value, namely $I(0) = \mathcal{E}_0/r$. We solve the equation by separation of variables, obtaining

$$I(t) = \frac{\mathcal{E}_0}{r} e^{-\frac{r+R}{L}t}. \quad (7.60)$$

We see that the current intensity exponentially decreases to zero with a time constant that is now $\tau' = L/(r + R)$, which is shorter than τ , much shorter if we take

$R \gg r$. Let us now suppose ourselves to be willing, starting from the steady regime with the switch in position 1, to take the current to zero by directly opening the switch. We might think the subsequent evolution to be similar to what we have just discussed with $R = \infty$ and that the current would decrease immediately. NEVER do that! Indeed, the current intensity would initially decrease very quickly. A very large induced emf would appear as a potential difference between the two sides of the switch. In practice, the circuit does not open because a spark develops through the switch. This phenomenon is momentous, especially if the inductance is large and can damage both the apparatus and the operator. To avoid that, configurations such as that in Fig. 7.12 or similar are used to supply a magnet. As we shall see in Chap. 8, energy is stored in a current-carrying coil, which is the energy of the magnetic field it creates. When the current is switched off, the field energy must be transferred somewhere else or, as in the case we considered, dissipated (meaning transferred to the microscopic level) on a resistor.

QUESTION Q 7.3. We connect an inductor of $L = 0.1 \text{ H}$ to a resistor with $R = 10 \Omega$ to a battery of negligible internal resistance. How much time after having closed the circuit will it take for the current intensity to reach $\frac{1}{2}$ of the steady value? \square

7.11 Alternating Current Circuits

Electromotive forces and the electric currents they produce may vary with time periodically or non-periodically or be constant. In Sect. 5.9, we studied electric circuits composed of resistors and constant emf generators. In Sects. 5.8 and 7.10, we studied currents and emfs that initially varied non-periodically to reach a stationary regime. We shall now discuss circuits in an alternating current regime. Let us start with a few definitions. *Direct current* (DC, for short) means a current that maintains a constant direction over time. The current intensity, however, may or may not vary. The corresponding emf maintains a constant polarity. Figure 7.13a shows an example of a direct, but not constant, current. *Constant current* means that both the direction and intensity of the current are constant over time, as shown in Fig. 7.13b. *Alternating current* (AC, for short), rigorously speaking, means a current, or voltage, that changes direction or polarity, respectively, over time. Figure 7.13c shows an example. The most common dependence on time of an AC is through a circular function, a sine or a cosine (which differ only by a $\pi/2$ phase difference), as in Fig. 7.13d. This is the case, for example, for the power distribution networks and the case that we shall now discuss. We shall follow the common use and talk simply of an alternating current, omitting the word sinusoidal.

The following discussion is completely analogous to the description we made in the third chapter of the 1st volume of the motions of the mechanical oscillators. We shall repeat the concepts here.

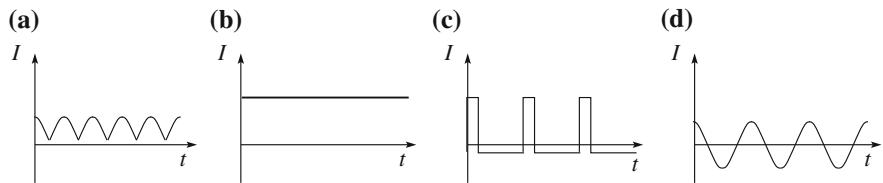


Fig. 7.13 Examples of **a** direct current, **b** constant current, **c** alternating current, **d** alternating sinusoidal current

We shall deal with electromotive forces with the following time dependence:

$$\mathcal{E} = \mathcal{E}_0 \cos(\omega t + \phi). \quad (7.61)$$

\mathcal{E}_0 is called the *amplitude*, the argument of the cosine, namely $\omega t + \phi$, is called the instantaneous *phase*, and ω the *angular frequency* or, sometimes, the *pulsation*. We see that the phase increases linearly with time starting from its initial value (namely at $t = 0$) ϕ , which is consequently called the *initial phase*. The angular frequency is linked to the *frequency* v and the *period* T by the well-known relations

$$\omega = 2\pi v = 2\pi/T. \quad (7.62)$$

Let us start with some general properties of the circular functions. Let x be any physical quantity varying with time as

$$x = A_0 \cos(\omega t + \phi). \quad (7.63)$$

It is often more convenient to work with exponential functions rather than circular ones. This is achieved by considering the x in Eq. (7.63) to be the real part of the complex quantity z , namely

$$z = A_0 \cos(\omega t + \phi) + i A_0 \sin(\omega t + \phi). \quad (7.64)$$

This expression can be written as follows:

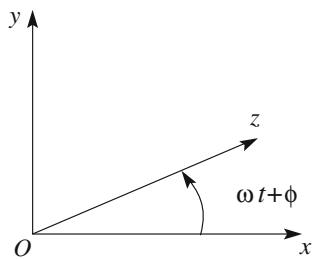
$$z(t) = A_0 e^{i(\omega t + \phi)} = A_0 e^{i\phi} e^{i\omega t} = z_0 e^{i\omega t}, \quad (7.65)$$

where, on the right-hand side, we have collected in z_0

$$z_0 = A_0 e^{i\phi}. \quad (7.66)$$

the time independent factors. The constant z_0 is called the *complex amplitude*. Its modulus is the oscillation amplitude, and its argument is the initial phase. The function $z(t)$ is the product of the (constant) complex amplitude and the factor $\exp(i\omega t)$, which gives the time dependence. The latter is a complex quantity of unitary

Fig. 7.14 Rotating vector representation



modulus and argument, varying linearly with time starting from a null initial value. Its rate of change is ω . Figure 7.14 represents graphically the function $z(t)$ in the complex plane, in which x and y are the real and the imaginary parts, respectively.

In Fig. 7.14, we have drawn a vector from the origin to the point $z(t)$. This vector represents our complex quantity. The vector rotates with time at a constant angular velocity equal to ω , starting from an angle ϕ with the x -axis. In the rotation, the vector keeps its magnitude constant, equal to the amplitude A_0 . We shall use this representation for several calculations. At the end, we shall get the physical quantity by taking the real part, namely

$$x(t) = \operatorname{Re}(z(t)) = \operatorname{Re}(A_0 e^{i\phi} e^{i\omega t}) = A_0 \cos(\omega t + \phi). \quad (7.67)$$

Geometrically, real part is the projection of the rotating vector on the x -axis

The complex notation of the harmonic oscillations is convenient for two main reasons: because exponentials are easier to handle than circular functions and because the derivative of an exponential is still an exponential. Note, however, that we can operate with a complex function and take the real part at the end if we perform only linear operations (sums, products by constants, differentiations and integrations), because they commute with the operation of taking the real part. We cannot follow the procedure when non-linear operations are involved, like taking the square of a circular function or the product of two of them.

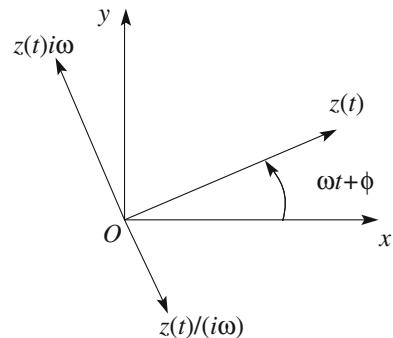
Moving forward, we shall need the derivative of $z(t)$, which is

$$\frac{dz(t)}{dt} = i\omega z_0 e^{i\omega t} = i\omega z(t). \quad (7.68)$$

We see that the derivative is equal to the original function multiplied by $i\omega$. Obviously, the physical dimensions of the derivative of z are the dimensions of z divided by a time. We can, however, represent $z(t)$ and $dz(t)/dt$ graphically in the same plane. In doing that, we recall that the imaginary unit can be expressed as $i = e^{i\pi/2}$. We can then write Eq. (7.68) in the form

$$\frac{dz(t)}{dt} = i\omega z_0 e^{i\omega t} = \omega z_0 e^{i\pi/2} e^{i\omega t} = \omega A_0 e^{i(\omega t + \phi + \pi/2)}. \quad (7.69)$$

Fig. 7.15 The function, its derivative and its primitive



We see that we can represent the derivative multiplying the function to differentiate by ω and adding $\pi/2$ to its phase. That is, the derivative is advanced in phase by a quarter of a period relative to the function, as shown in Fig. 7.15.

When we take the real part, we obtain

$$x(t) = \operatorname{Re} \frac{dz(t)}{dt} = \omega A_0 \cos(\omega t + \phi + \pi/2) = -\omega A_0 \sin(\omega t + \phi), \quad (7.70)$$

as we expected.

Let us now look at the primitive of $z(t)$. We easily find that

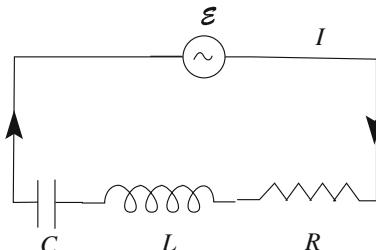
$$\int z(t) dt = \frac{z(t)}{i\omega} = -i \frac{z(t)}{\omega}. \quad (7.71)$$

We see that the primitive is obtained simply by dividing the function by ω and subtracting $\pi/2$ from its phase. The primitive is delayed in phase by a quarter of a period relative to the function, as shown in Fig. 7.15.

We now go back to the physical problem. We shall consider electric circuits, including the following elements: alternate (sinusoidal) electromotive force generators (angular frequency ω), resistors, capacitors and inductances. These are called *passive components*, while the emf generators are active.

In Fig. 7.16, all the elements are in series with their symbols. The elements are an alternate emf generator, a resistor of resistance R , an inductor of inductance

Fig. 7.16 An RLC circuit in series



L and a capacitor (or condenser) of capacitance C . To find the equation governing this circuit, let us consider having a unit charge going about the full circuit. Its final energy is equal to the initial one. We shall then express in a formula the fact that the sum of the work done on the charge is zero.

The unit charge gains the energy \mathcal{E} crossing the generator. It loses the energy $RI(t)$ crossing the resistor. When crossing the inductor, the charge sees the induced emf $-LdI/dt$. Consequently, the lost energy is $+LdI/dt$. The charge does not cross the capacitor physically, but things proceed as if it did, because if $Q(t)$ is the charge on one plate, the charge of the other one is $-Q(t)$. Now, $Q(t)$ varies with time, because the current brings charges on the first plate. The current intensity is the arriving charge per unit time, namely

$$I(t) = \frac{dQ(t)}{dt}. \quad (7.72)$$

For what we just stated, the current intensity arriving at the second electrode is always $-I$. But this current has a direction opposed to the positive orientation of the circuit. Consequently, the current leaving the second plate is, again, I . Things proceed as if our unitary charge would go from one electrode to the other. In doing that, the unit charge loses an energy equal to $Q(t)/C$.

We can now write the energy balance equation as

$$\mathcal{E}(t) = RI(t) + L\frac{dI(t)}{dt} + \frac{Q(t)}{C}. \quad (7.73)$$

Notice that the three terms on the right-hand side are the electromotive forces through the three elements of the circuit, or, as is often improperly stated, the potential drops. As such, Eq. (7.73) is very similar to Ohm's law. Equation (7.73) is a differential equation in the unknown function $I(t)$. In order to see that explicitly, we differentiate the equation, finding

$$L\frac{d^2I(t)}{dt^2} + R\frac{dI(t)}{dt} + \frac{I(t)}{C} = \frac{d\mathcal{E}(t)}{dt}. \quad (7.74)$$

This is a non-homogeneous differential equation of the second order with constant coefficients. We shall find its solution in the next section.

7.12 Complex Impedance

Let us find the solution to Eq. (7.74). We can always choose the origin of time in order to have the initial phase of the electromotive force be equal to zero, namely

$$\mathcal{E} = \mathcal{E}_0 \cos \omega t. \quad (7.75)$$

We write it in a complex notation as

$$\mathcal{E} = \mathcal{E}_0 e^{i\omega t}. \quad (7.76)$$

We search for a solution to Eq. (7.74) in the form

$$I = I_0 e^{i(\omega t + \phi)} \quad (7.77)$$

with amplitude and initial phase to be determined. Differentiating as we did in Eq. (7.74), we immediately have

$$i\omega \mathcal{E}_0 e^{i\omega t} = (i\omega R - \omega^2 L + 1/C) I_0 e^{i(\omega t + \phi)}.$$

We see that all the terms have the same time dependence, which consequently can be simplified out, obtaining

$$i\omega \mathcal{E}_0 = (i\omega R - \omega^2 L + 1/C) I_0 e^{i\phi}.$$

This is now an algebraic equation in the unknowns I_0 and ϕ . In order to obtain an expression similar to Ohm's law, let us divide all the terms by $i\omega$ and multiply them back by $\exp(i\omega t)$. We obtain

$$\mathcal{E} = (R + i\omega L - i/(\omega C)) I. \quad (7.78)$$

We can say the emf provided by the generator is equal to the sum of the drops across each of the elements in series. The drop on each element is equal to the current intensity times a coefficient characteristic of the element, called the *impedance* of the element. The impedance of a resistor is equal to its resistance, and is called the "ohmic impedance", and is expressed as

$$Z_R = R. \quad (7.79)$$

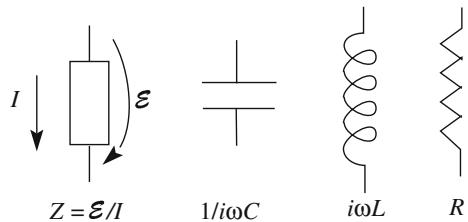
The impedance of an inductor, called the *inductive impedance*, is the imaginary quantity

$$Z_L = i\omega L. \quad (7.80)$$

The impedance of a capacitance, called the *capacitive impedance*, is the imaginary quantity

$$Z_C = -i \frac{1}{\omega C}, \quad (7.81)$$

Fig. 7.17 The basic AC circuit elements



All these quantities are dimensionally homogeneous and are measured in ohm (Ω). In this notation, Eq. (7.78) becomes

$$\mathcal{E} = (Z_R + Z_L + Z_C)I. \quad (7.82)$$

Figure 7.17 summarizes these results graphically. It gives, for each element, its impedance, namely the ratio between voltage drop between its terminals and the current intensity crossing the element.

As opposed to the resistance, both the inductive and capacitive impedances depend on angular frequency. The inductive impedance grows proportionally to the angular frequency. Indeed, the larger the rate of change of the magnetic flux linked to the inductor, the larger the induced emf. The capacitive impedance decreases in inverse proportion to the angular frequency. Indeed, the impedance of a capacitor is infinite in the limit of constant current. The mechanism we have described allows for an equivalent passage of current through the capacitor, but only for currents variable with time.

To get an idea of the orders of magnitude, consider the network frequency of 50 Hz (namely $\omega = 314$ Hz). The impedance of an inductor, for example, of 10 mH, is about 3Ω . It becomes about $300 \text{ k}\Omega$ at 5 MHz frequency. The impedance of a $1 \mu\text{F}$ capacitor is about $3 \text{ k}\Omega$ at the network frequency and only $30 \text{ m}\Omega$ at 5 MHz.

The inductive and capacitive impedances are imaginary numbers. What does this mean? Remember that the impedance of a circuit element is, by definition, the ratio between the voltage drop across it and the current intensity through it. Both these quantities are complex in the notation we are using. As we have already seen, multiplying by the imaginary unit i means advancing the phase by $\pi/2$, while dividing by i means retarding the phase by $\pi/2$. Physically, the emf between the terminals of a resistor is in phase with the current, the emf between the terminals of an inductor is in phase advance of $\pi/2$ to the current, and the emf between the plates of a capacitor is in phase delay of $\pi/2$ to the current.

Equation (7.78) can be written in an even more compact form as

$$\mathcal{E} = ZI. \quad (7.83)$$

where the complex quantity Z is called the *total impedance* of the circuit. As with any complex quantity, we can write Z in two ways. The first is as sum of a real and an imaginary part (the latter is sometimes called *reactance*), namely

$$Z = R + i[\omega L - 1/(\omega C)] \equiv R + iX. \quad (7.84)$$

The second is in terms of its modulus and its argument. Let us call them Z_0 and $-\phi$, respectively (we shall soon see the reason for the minus sign). We can write

$$Z = Z_0 e^{-i\phi}. \quad (7.85)$$

Multiplying the complex quantity I by Z , as in Eq. (7.83), means multiplying the modulus of I by Z_0 and diminishing its argument by ϕ .

Figure 7.18 shows the different terms of Eq. (7.84) and their sum. We immediately find the values of I_0 and ϕ by adding the vectors. We have

$$I_0 = \frac{\mathcal{E}_0}{\sqrt{R^2 + [\omega L - 1/(\omega C)]^2}} = \frac{\mathcal{E}_0}{Z_0}, \quad (7.86)$$

where

$$Z_0 = \sqrt{R^2 + [\omega L - 1/(\omega C)]^2}. \quad (7.87)$$

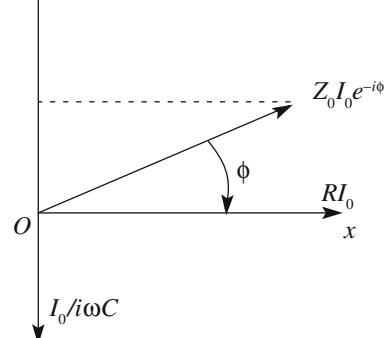
is the modulus of the total complex impedance of the circuit, and

$$\tan \phi = \frac{1/(\omega C) - \omega L}{R}. \quad (7.88)$$

where ϕ is the opposite of the complex impedance argument.

We notice that the impedance of our circuit is very small for a particular value of the angular frequency, namely when $\omega L = 1/(\omega C)$, that is for $\omega = (LC)^{-1/2}$. If the resistance R is small, the total impedance may be close to zero. Under these conditions, the inductive and capacitive impedances, which are always in phase

Fig. 7.18 Rotating vector representation of the electromotive forces in a RCL circuit



opposition to one another, have the same modulus and cancel one another out. This is the resonance phenomenon, completely similar to the mechanical one we studied in Sect. 3.9 of the 1st volume. Resonance is a common phenomenon in nature. We shall discuss it completely in Volume 4.

QUESTION Q 7.4. An AC circuit includes a resistor, an inductor and a capacitor. What are the phase differences (value and sign) between the emf and current intensity in each element? □

QUESTION Q 7.5. By what factor does the impedance of a resistor vary when the frequency changes from 50 to 100 Hz? And that of an inductor? And that of a capacitor? □

Let us now discuss a few assumptions we have made in our analysis. We first notice that we have assumed the current intensity in a given instant of time to be exactly the same in all the sections of the circuit. In practice, the intensity variations propagate with a very high, but not infinite, velocity. The propagation velocity depends on the circuit materials, but is always on the order of the speed of light. For a typical geometrical dimension of a circuit, say 10 cm, the time for a current variation to cross is on the order of 1 ns. The hypothesis of independence of the current intensity of the position along the circuit is then valid if the oscillation period is much larger than that. We can say that the oscillation frequency should not exceed 100 MHz or so.

The circuit elements are physical objects, rather than the ideal elements appearing in the equations. Let us take a closer look at them.

An inductor is made by wrapping a long wire in many loops, often around a nucleus (think of a small torus, for example) of ferromagnetic material, to increase the self-inductance, as we shall see in Sect. 9.8. The wire always has a resistance, which is not always negligible. In addition, each pair of adjacent loops acts as a small capacitor with an electric field between the two loops. Consequently, the circuit element that we call an inductor has a capacitance as well. In order to provide a more precise picture of a physical inductor, we can represent each pair of loops as an inductor and a resistor in series, in parallel with a capacitor, as shown in Fig. 7.19.

The relative importance of the sub-elements depends on the frequency. If the frequency is very low, the inductor behaves very similarly to a resistor, exactly like a resistor in the continuous current limit. When the frequency is sufficiently high, the impedance of the inductor element is much larger than that of the resistor and much smaller than that of the capacitor in parallel. Consequently, the element tends to behave as an ideal inductor. At even higher frequencies, the impedance of the capacitor becomes much smaller than that of the inductor and the current “prefers”

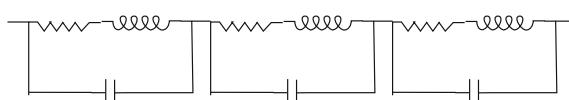


Fig. 7.19 Detailed diagram of a real inductor

going through the capacitors, jumping from one loop to the next, rather than flowing all through the loop.

The real capacitors behave quite similarly to the ideal ones at low frequencies and when they have a vacuum between the plates, a situation that rarely happens in practice. In practical cases, a dielectric is always present. Dielectric materials are very good insulators, but not perfect. Even if small, some current gets through them. We can represent the capacitor including a high resistance resistor in parallel. In addition, in the ideal capacitor, the electric field is enclosed between the plates and there is no magnetic field. Contrastingly, as we shall see in Chap. 10, any electric field variable with time generates a magnetic field, whose lines are around the capacitor in this case. The effect is that much larger the higher the change rate of the electric field, namely the frequency of the AC current. We must be aware that the behavior at very high frequencies of all real capacitors is very different from the ideal ones.

The resistance of an ideal resistor is independent of frequency. This is not true for the real ones. In practice, the current flowing in the resistor produces a magnetic field, which varies with time because so does the current. The electric field induced by the variable magnetic field produces a decrease in the current density near the axis of the resistor. The current tends to flow in a layer close to the surface, the thickness of which decreases with increasing frequency. This is called the skin effect. The consequence is that the resistance of the resistor decreases with increasing frequency. If, as is often the case, the resistor is made of a long wire, it also has an inductance (in series) and a capacitance (in parallel).

In Sect. 5.9, we studied how to analyze circuits composed of several loops of resistors and DC voltage generators, using Ohm's law and the Kirchhoff rules that are consequences of energy and charge conservations. Now, in the AC regime, we have found Eq. (7.82), which is formally identical to Ohm's law. Consequently, to analyze complex circuits, we can use the same rules, the rule of the nodes and the rule of the loops, with the only differences being that we are now working with complex quantities and the relative phases matter. The loops rule is

$$\sum_{i=1}^n Z_i I_i = \sum_{i=1}^n \mathcal{E}_i. \quad (7.89)$$

As we have assumed a steady regime and the current intensity to be independent of the position along the circuit, it is still true that the algebraic sum of the currents entering a node is zero. The nodes rule is

$$\sum_{k=1}^n I_k = 0. \quad (7.90)$$

7.13 Energy Balance in a Circuit

We continue to consider the circuit in Fig. 7.16. The power delivered by the generator, namely the electric work per unit time, is

$$W = \mathcal{E}I = \mathcal{E}_0 \cos(\omega t)I_0 \cos(\omega t + \phi), \quad (7.91)$$

which obviously varies with time. Let us start by considering the capacitor. We know that the stored energy is $Q^2(t)/(2C)$. This energy varies periodically in time from a minimum, which is zero, to a maximum. The variation does not correspond to losses or gains of energy, but rather to transfers of energy in other components of the circuit. The situation of the inductor is similar. As we shall learn in the next chapter, the stored energy is $I^2(t)L/2$, which varies periodically as well, transferred back and forth between other parts of the circuit. As a matter of fact, the generator must at any time provide the electric work to charge and discharge the capacitor and the inductor. This work is sometimes positive and sometimes negative, averaging to zero in every period. Contrastingly, the electric work delivered by the generator on the resistor is dissipated. This is given by Eq. (7.91), which can be written as

$$W = \mathcal{E}_0 I_0 \cos\phi \cos^2\omega t - \mathcal{E}_0 I_0 \sin\phi \cos\omega t \sin\omega t. \quad (7.92)$$

Let us calculate the mean value over a period, recalling that the average on a period of a circular function squared ($\cos^2\omega t$) is $\frac{1}{2}$. We define the *effective values* of the electromotive force and the current intensity as the square roots of the mean values over a period of their squares. The effective values are equal to the maximum values divided by $\sqrt{2}$, namely

$$\mathcal{E}_e = \mathcal{E}_0/\sqrt{2}; \quad I_e = I_0/\sqrt{2}. \quad (7.93)$$

We obtain

$$\langle W \rangle = \frac{\mathcal{E}_0 I_0}{2} \cos\phi = \mathcal{E}_e I_e \cos\phi. \quad (7.94)$$

This expression is similar to that which we found for constant currents. However, not only do we now have the product of the effective values of the emf and current intensity, but that of the cosine of the relative phase as well. From Eq. (7.86), we have $\mathcal{E}_0 = Z_0 I_0$, and from Eqs. (7.84) and (7.85), we get

$$R = Z_0 \cos\phi. \quad (7.95)$$

We can conclude that

$$\langle W \rangle = \frac{Z_0 I_0^2}{2} \cos\phi = \frac{1}{2} R I_0^2 = R I_e^2, \quad (7.96)$$

which directly shows that the mean power delivered by the generator is steadily dissipated on the resistor. From Eq. (7.95), we learn that $\cos\phi$ cannot be negative and that, consequently, $-\pi/2 \leq \phi \leq +\pi/2$.

Let us finally notice that, in this section, we did not use the complex notation. We could not use it, because we performed non-linear operations, like the product of two circular functions or their square, which do not commute with the real part taking.

Summary

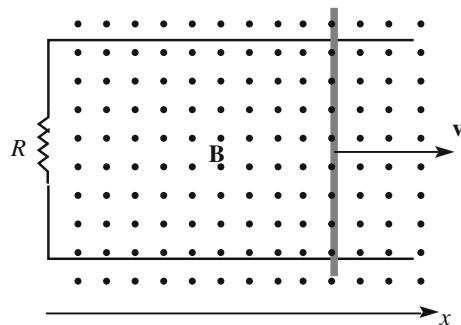
In this chapter, we began the study of time-dependent electromagnetic phenomena. We discussed electromagnetic induction, learning the following principal concepts:

1. The flux rule, its two different origins and its exceptions
2. That the electric field is not conservative (and not dissipative) under dynamic conditions
3. Faraday's law
4. The exceptions to the flux rule
5. The eddy currents
6. The mutual induction and the self-induction
7. The behavior of relevant circuits in a transient regime
8. The behavior of relevant circuits in a stationary sinusoidal regime
9. The complex impedance of the different circuit elements.

Problems

- 7.1. Can the electric field lines be closed? Can they radiate from a point?
- 7.2. A circular circuit is immersed in a uniform magnetic field \mathbf{B} directed normally to the circuit towards the observer. What is the direction of the current if (a) the field increases with time, (b) it decreases, (c) the circuit is expanded, or (d) it is contracted?
- 7.3. Does the inductance of a circuit depend or not on: the electric resistance of the circuit, the current intensity in the circuit, the intensity of the magnetic field in which it might be immersed, the form and the dimensions of the circuit, the material of the circuit, the possible presence of other circuits?
- 7.4. A wire of length $l = 5$ cm moves in a uniform and constant magnetic field $B = 1.6$ T with speed $v = 10$ m/s. What is the emf at the ends of the wire in

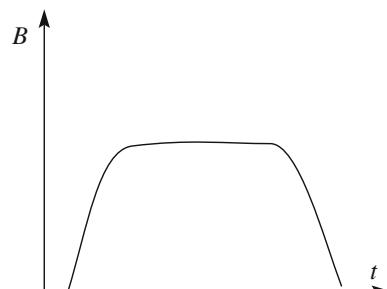
Fig. 7.20 Bar moving with constant velocity in a magnetic field



the following cases: (a) the velocity is parallel to the wire and normal to the field, (b) the velocity is normal to the wire and parallel to the field, (c) velocity, wire and field are perpendicular to one another.

- 7.5. Consider Fig. 7.20. The bar moves to the right at constant speed v . If Q is the heat dissipated by the induced current per unit time, what is the force F acting on the moving bar?
- 7.6. An electric charge crosses a space point at a certain instant with a certain kinetic energy. Under the action of electromagnetic forces, it later passes again through the same point with twice as much kinetic energy. Is this possible?
- 7.7. Does the mutual inductance coefficient depend or not on: the electric resistance of the circuits, the current intensities they carry, the magnetic field intensity in which they are located, the shapes and sizes of the circuits, the materials they are made of, the presence of other circuits, the relative positions of the two circuits?
- 7.8. We want to insert an aluminum plate between the poles of a powerful electromagnet producing a field of 3 T. Is it better to make that slowly or quickly?
- 7.9. A coil is in a magnetic field. The field is normal to the coil and its magnitude varies with time, as shown in Fig. 7.21. Make a diagram showing the evolution of the induced current.

Fig. 7.21 Magnetic field intensity versus time through the coil of problem 7.9



- 7.10. A circuit has the form of a closed plane curve carrying a current of known intensity. We try to determine its self-inductance by measuring the magnetic field intensity at a number of points of a plane bounded by the curve. Are these measurements sufficient?
- 7.11. Two equal inductors can be connected in series or in parallel. In which case is the total inductance larger?
- 7.12. The mean field $\langle B \rangle$ of a betatron of radius $R = 0.3$ m increases linearly in time in 1 ms from zero to 0.3 T. How many turns must the electrons do before reaching the kinetic energy of 10 MeV? What is the distance they travel?
- 7.13. Can a circuit of null reactance be created? If yes, how?
- 7.14. A circuit is made of a resistor $R = 1$ k Ω , an inductor $L = 280$ mH and a capacitor $C = 8$ μF in series fed with an alternate current at 50 Hz frequency and 220 V effective voltage. Find: (a) the effective value of the current, (b) the phase difference between the emf between the extremes of the series and the current, (c) the effective values of the emf at each of the elements.
- 7.15. An electric oven, whose heating elements have both a resistance and an inductance, is powered from the power network at a voltage of effective value of 220 V. The effective value of the current is 5 A. With these data, can we state that the oven absorbs 1.1 kW?
- 7.16. How could you make a circular electric field line?
- 7.17. An airplane flies horizontally at 900 km/h in a region in which the vertical component of the earth's field is 20 μT . Find the emf between the tips of the wings if their distance is 25 m.

Chapter 8

Magnetic Energy

Abstract In this chapter, we study the energy of systems of steady currents, namely under conditions independent of time. The current-carrying circuits store energy, which is proportional to the square of the current intensity and is greater the larger the inductances. Currents generate magnetic fields, and we can think of the energy of the system as energy stored in that field as well. Energy is distributed throughout the entire space with a density proportional to the square of the magnetic field.

In this chapter, we study the energy of system of steady currents, namely under conditions independent of time. In Chap. 3, we made a similar study in electrostatics. The current-carrying circuits store energy, which increases as the square of the current intensity and is greater the larger the inductances. A circuit of large self-inductance carrying an intense current, such as, typically, those of the big electromagnets, may store a very remarkable quantity of energy. This can be dangerous if not properly managed.

We shall start in Sect. 8.1 with the study of a single circuit carrying a steady current and continue in the subsequent section by studying the case of more than one circuit. We shall see that the total energy is the sum of the energies proper of each circuit, separately considered, and of interaction terms, one for each pair of circuits. In Sect. 8.3, we consider a particular pair of circuits, namely an elementary loop, which is a magnetic dipole, and the circuit producing a steady magnetic field in which the dipole is immersed. We further develop the concepts already discussed in Sect. 6.14.

Finally, in Sect. 8.4, we shall see that, similarly to the case of the electric field, magnetic energy can be thought to be stored in the magnetic field. It is distributed throughout the entire space with a density proportional to the square of the magnetic field. We shall show that the expressions of energy as a property of the currents or as stored in the magnetic field are equivalent under time-independent conditions. We shall finally anticipate that, under dynamic conditions, only the expression of the magnetic field energy remains valid.

8.1 Energy of a Steady Current

Let us consider a circuit at rest in a vacuum carrying a steady current of intensity I , and let us ask ourselves what the energy of the system is. The energy is equal to the (electric) work required to “switch on” the current, bringing its intensity from zero to I (an amount we get back returning the current to zero).

Every circuit has a non-zero inductance, because there is always a magnetic flux linked to the circuit. Let us call it L and schematize it as a lumped element, namely an inductor, as shown in Fig. 8.1. To obtain the current, we need an emf generator. Let us assume it to be a battery of constant emf \mathcal{E}_0 . Let R be the total resistance of the circuit and let us schematize it again as a lumped element, namely a resistor. When the switch Sw is closed and the steady regime has been reached, the current intensity is $I = \mathcal{E}_0/R$. Under these conditions, the current intensity does not vary with time and the inductor does not act.

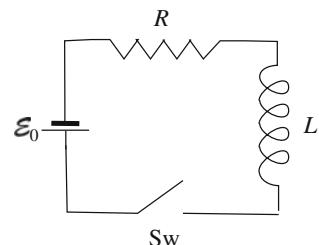
Let us start from the initial state in which the switch is open and the current is zero. We close the switch and calculate the work done by the generator to bring the current to its steady value. We can always assume that the current varies slowly enough for its intensity to be equal in all the sections of the circuit. Let us call i the instantaneous current intensity during the process. As i increases, a back electromotive force $\mathcal{E}_L = -Ldi/dt$ develops. The generator must spend work per unit time equal to $-\mathcal{E}_L i = Lidi/dt$ against it. This corresponds to the work in the time interval dt equal to $dw = Lidi$. The total work to change i from zero to I is then

$$W = \int_0^I Lidi = \frac{1}{2}LI^2, \quad (8.1)$$

which is the energy of the system. We can call it the *energy stored in the inductor* and also the *energy of the current*. As we shall see, this is the energy of the magnetic field generated by the current I .

Notice that the only role of the resistor R in our argument is to limit the steady current at the value $I = \mathcal{E}_0/R$. Obviously, due to the presence of R , the generator must spend the electric power Ri , which is dissipated via the Joule effect. This fact,

Fig. 8.1 The simplest circuit including an inductor



however, has nothing to do with the energy of the circuit. The dissipated electrical power cannot be taken back. Contrastingly, we can get back the energy $\frac{1}{2}LI^2$, as mechanical or electrical work.

8.2 Energy of a System of Steady Currents

Let us now consider a system of n circuits at rest in a vacuum having inductances L_1, L_2, \dots, L_n , carrying the steady currents of intensities I_1, I_2, \dots, I_n , and let us calculate the energy of the system. As always, this energy is the work to be spent to build the system, namely to bring the current intensities from zero to their regime values. This is also the work we can get back extinguishing the currents. Figure 8.2 shows the case of four currents.

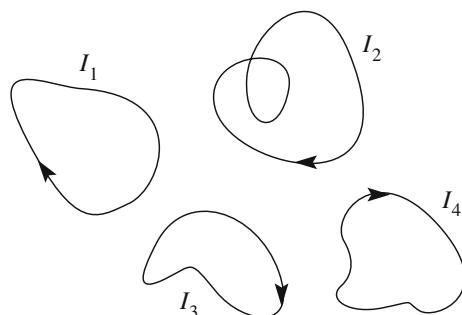
Initially, all the currents are zero. Let us switch on the current in circuit 1. This is the case considered in the previous section. The electric work to be spent is $L_1I_1^2/2$. Next, we switch on the current in circuit 2. Again, to bring it to its steady value I_2 , the generator of circuit 2 must spend the electrical work $L_2I_2^2/2$. However, there is something else now. Indeed, when the current in circuit 2, which we call i_2 , increases, the magnetic flux it generates linked to circuit 1 varies in time, producing an emf $\mathcal{E}_{21} = -M_{21}di_2/dt$ in that circuit. M_{21} is the mutual inductance coefficient between the two circuits (which we have oriented). The work spent by the generator in circuit 1 in dt is $-\mathcal{E}_{21}I_1dt = +M_{21}I_1di_2$. The total work is found integrating on i_2 from 0 to I_2 . We have

$$\int_0^{I_2} M_{21}I_1di_2 = M_{21}I_1I_2.$$

Summing up, the energy of circuits 1 and 2 is

$$U_{21} = \frac{1}{2}L_1I_1^2 + \frac{1}{2}L_2I_2^2 + M_{21}I_1I_2.$$

Fig. 8.2 A system of steady currents



Going directly to n circuits, the total energy of the system is

$$U = \frac{1}{2} \sum_{k=1}^n L_k I_k^2 + \frac{1}{2} \sum_{h=1}^n \sum_{k=1, k \neq h}^b M_{hk} I_h I_k. \quad (8.2)$$

The first sum is the sum of the proper energies of each circuit; the second sum is the sum of the interaction energies of each pair of circuits. The factor $\frac{1}{2}$ for the second sum is there to cancel out the double counting resulting from having counted the same pair in the double sum as h, k and k, h .

Note that, while the proper energies are always positive, the interaction energies may be positive or negative.

8.3 Energy of a Dipole

In this section, we shall discuss the energy of a magnetic dipole in a steady magnetic field \mathbf{B} . As a matter of fact, the problem is the same as that which we discussed in the last section, seen from a somewhat different point of view. Indeed, the dipole is just a loop carrying a steady current I_1 , and we can think of the field \mathbf{B} as having been produced by a second circuit. To fix the ideas, let us consider it to be a coil and let us call it circuit 2. Let I_2 be the steady current it carries. To be precise, what we are looking for is the *interaction energy* between dipole and field, namely between dipole and coil. This energy can be thought of as the work to be done starting from an initial state in which the two circuits already carry their currents but do not interact. We can consider initially having the loop very far from the coil, where the field of the coil is practically zero. The interaction energy is the work needed to bring the loop to its final position, while the currents in the loop and the coil keep their values constant. As a matter of fact, we already know the answer. If M is the mutual inductance between loop and coil, their interaction energy is

$$U_{\text{int}} = MI_1 I_2. \quad (8.3)$$

Note that if we call Φ_{21} the flux generated by the coil (circuit 2) linked to the dipole (circuit 1), then $\Phi_{21} = MI_2$, and we can write Eq. (8.3) as

$$U_{\text{int}} = \Phi_{21} I_1. \quad (8.4)$$

On the other hand, the loop carrying the current I_1 generates a magnetic field as well. The flux of this field linked to the coil is $\Phi_{12} = MI_1$ with the same M . We can then write Eq. (8.3) in yet another form, namely

$$U_{\text{int}} = \Phi_{12} I_2. \quad (8.5)$$

The three equations express the same quantity, the interaction energy.

It is instructive to calculate the interaction energy in still another way, by evaluating the mechanical work in the process we mentioned above, namely the work needed to move the dipole from outside to inside the field. We shall find an expression of U_{int} in terms of the dipole moment μ and of the field \mathbf{B} . Let us simplify the issue, assuming the field \mathbf{B} everywhere to have the same direction parallel to and in the same sense as μ and to depend on x alone (independent of y and of z). Let us assume the loop to be rectangular, with side lengths Δx and Δy along the two corresponding directions.

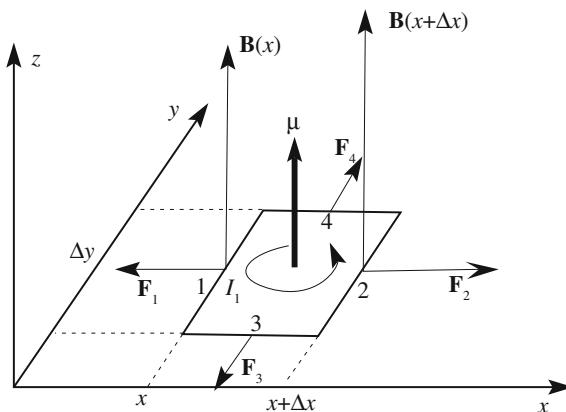
The loop is initially far away, on the side of the negative x , where the field is zero. Let us calculate the mechanical work against the field forces required to move the coil into its final position, which we call x_F . Figure 8.3 shows the relevant quantities. Four forces resulting from the magnetic field act on the loop, one on each of its sides. The forces are in the xy plane and perpendicular to the length of each side. Two forces, \mathbf{F}_3 and \mathbf{F}_4 , are normal to the displacement and do not produce any work. The work done for the infinitesimal displacement dx by the other two forces, \mathbf{F}_1 and \mathbf{F}_2 , is

$$\begin{aligned} dW &= -(F_2 - F_1)dx = -[I_1 \Delta y B(x + \Delta x) - I_1 \Delta y B(x)]dx \\ &= -I_1 \Delta y [B(x + \Delta x) - B(x)]dx. \end{aligned}$$

Now, if, as we assume, the geometrical dimensions of the dipole are small compared to the distance on which \mathbf{B} varies appreciably, we can write the above expression as

$$dW = -I_1 \Delta y \Delta x \frac{dB}{dx} dx = -\mu \frac{dB}{dx} dx$$

Fig. 8.3 A magnetic dipole during its translation inside the field



The mechanical work needed to bring the dipole into its final position is then

$$W_{\text{mec}} = -\mu \int_{-\infty}^{x_F} \frac{dB}{dx} dx = -\mu \int_{B(-\infty)}^{B(x_F)} dB = -\mu B(x_F)$$

In other words, the *mechanical energy* we have put into the system is

$$U_{\text{mec}} = -\boldsymbol{\mu} \cdot \mathbf{B}. \quad (8.6)$$

We have written the right-hand side as the dot product of two (axial) vectors in order to have an expression independent of the reference frame. This result, which we obtained considering a particular field shape, is, in fact, valid in general. Let us recall that we already found the dipole mechanical energy in Eq. (6.63).

In the case we are considering, as one sees in Fig. 8.3, the vectors $\boldsymbol{\mu}$ and \mathbf{B} have the same direction and sense, and U_{int} is negative. Indeed, while we move the dipole toward regions of higher field, namely with $F_1 > F_2$, the field forces attract the loop inside. In other words, the work done against the field forces is negative. The statement is true, however, as we shall see immediately, if both magnetic moment and field intensity are kept constant during the operation. This is not something that one can just assume, because during the translation, both current intensities, in the coil and in the loop, tend to vary. To keep both constant, two *electric* works are necessary, which we now calculate.

When the loop moves toward the higher field regions, the linked flux Φ_{21} grows, inducing an emf that tends to vary the current I_1 in the coil. The emf is

$$\mathcal{E}_1 = -\frac{d\Phi_{21}}{dt}. \quad (8.7)$$

If v_t is the velocity, the loop moves by $dx = v_t dt$ in the time interval dt . Correspondingly, the linked flux grows by $B(x + \Delta x)dx\Delta y$ on side 2, and diminishes by $B(x)dx\Delta y$ on side 1. In total, the flux variation is

$$d\Phi_{21} = B(x + \Delta x)dx\Delta y - B(x)dx\Delta y = [B(x + \Delta x) - B(x)]dx\Delta y = \frac{dB}{dx}\Delta x\Delta y dx$$

To maintain the current I_1 as constant, it is necessary to have a generator in the loop delivering the electric power $-\mathcal{E}_1 I_1$, hence the (electric) work in the time interval dt

$$dW_{\text{el-loop}} = -\mathcal{E}_1 I_1 dt = I_1 d\Phi_{21}. \quad (8.8)$$

Using the above-found result, this can be written as

$$dW_{\text{el-loop}} = (I_1 \Delta x \Delta y) \frac{dB}{dx} dx = \mu dB. \quad (8.9)$$

Remembering that field and magnetic moment have the same positive direction, and using Eq. (8.8), the total electric work of the generator in the loop is then

$$U_{\text{el-loop}} = +\boldsymbol{\mu} \cdot \mathbf{B} = I_1 \Phi_{21}. \quad (8.10)$$

We have reached an apparently surprising result. The total work on the loop, the sum of the mechanical and the electrical, is zero. Looking more closely, however, the result could have been foreseen from the start. Let us consider a charge carrier. Its velocity is just one, the vector sum of its drift velocity in the wire and the (macroscopic) velocity of the wire itself. The work of the magnetic force is zero, because it is normal to the carrier velocity. We have calculated the mechanical and electrical works considering the two parts of the carrier velocity separately and found two equal and opposite quantities, adding up to zero. Let us look at the issue in detail.

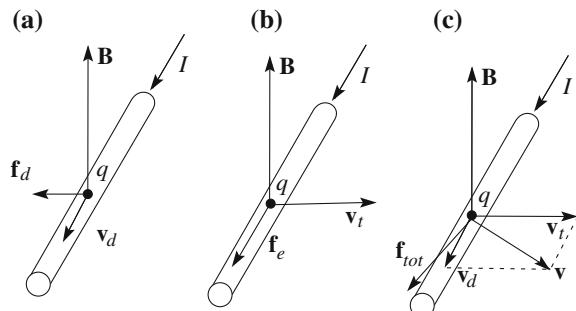
To be concrete, suppose the carriers have positive charge q . Let us consider side 1. The carriers move through the wire with an average speed, the drift velocity \mathbf{v}_d , which is parallel to the wire. The corresponding magnetic force is $\mathbf{f}_d = q\mathbf{v}_d \times \mathbf{B}$ (see Fig. 8.4a). The force \mathbf{F}_1 acting on side 1, which we considered in calculating the mechanical work, is the resultant of these forces on all the carriers.

We have calculated the electric work considering the motion of the wire that carries all the carriers with velocity \mathbf{v}_t . The corresponding magnetic force on the single carrier is $\mathbf{f}_e = q\mathbf{v}_t \times \mathbf{B}$, which is parallel to the wire (see Fig. 8.4b). The resultant of these forces on all the carriers is the origin of the electromotive force.

Indeed, the carrier has a single velocity, which is $\mathbf{v} = \mathbf{v}_d + \mathbf{v}_t$. The total magnetic force $\mathbf{f} = q\mathbf{v} \times \mathbf{B}$, being normal to \mathbf{v} , does not make any work (see Fig. 8.4c). The sum of the mechanical and electrical works on the loop is just the sum of these null works.

We have still to calculate the electrical work that is needed to maintain the magnetic field \mathbf{B} as constant during the motion of the loop. This work is done by the

Fig. 8.4 Forces on a positive carrier in a translating wire, corresponding to **a** drift velocity, **b** translation velocity, **c** total velocity



generator that produces the current I_2 in the coil. The loop produces a magnetic field, whose flux linked to the coil is Φ_{12} . The flux varies when the loop moves, inducing in the coil the emf $\mathcal{E}_2 = -d\Phi_{12}/dt$. To keep I_2 constant, the generator in circuit 2 must deliver the electric power $-I_2\mathcal{E}_2$, namely the (electrical) work in the time interval dt

$$dW_{\text{el-coil}} = -\mathcal{E}_2 I_2 dt = I_2 d\Phi_{12}. \quad (8.11)$$

The total electrical work on the coil for its complete displacement, using Eqs. (8.3), (8.4) and (8.10), is then

$$U_{\text{el-coil}} = I_2 \Phi_{12} = +\boldsymbol{\mu} \cdot \mathbf{B}. \quad (8.12)$$

We see that $U_{\text{el-coil}}$ is identical to $U_{\text{el-loop}}$, namely the electrical work made by each of the generators in the two circuits is identical. This result is not surprising. Indeed, the motion of the loop is a relative motion of the two circuits. An observer in a reference frame linked to the loop sees the loop at rest and the coil moving. The roles of the two circuits are inverted, but obviously, the work done by each of them does not change.

In conclusion, the total interaction energy is given by the sum of the three terms we have found, the mechanical energy and the two electric energies. The interaction energy is

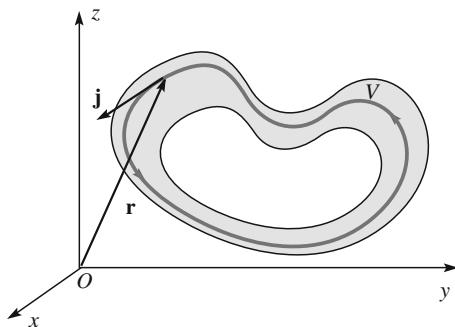
$$U_{\text{tot}} = +\boldsymbol{\mu} \cdot \mathbf{B}. \quad (8.13)$$

This is equal to the opposite of U_{mec} . As we have already noticed, the two circuits attract one another in the situation we are discussing in which $\boldsymbol{\mu}$ and \mathbf{B} have the same direction and sense, provided that the current intensities in the two circuits are kept constant. The system tends to evolve toward the state of minimum mechanical energy. Similarly, if we consider a dipole, namely our elementary loop, free to turn in a magnetic field, its stable equilibrium position is the position of minimum mechanical energy, namely in which $\boldsymbol{\mu}$ and \mathbf{B} have the same direction and sense only if the system develops at constant $\boldsymbol{\mu}$ and \mathbf{B} .

8.4 Energy of the Magnetic Field

In Sect. 3.5, we saw that an electric field contains energy, which is distributed in space with a density proportional to the square of the field. We shall now see that the energy associated with the electric currents that we just discussed can similarly be thought of as energy in the magnetic field, with a density proportional to the square of its magnitude. The two situations are similar, but there are differences as well. Indeed, in the case of the electric field, the energy of the system of charges is equal to the work to be done from outside to assemble the system against the

Fig. 8.5 A region of volume V containing a system of steady electric currents



electric forces exerted by the charges we are moving, which are the forces of their electric field. In the magnetic case, the work of magnetic forces is zero.

Consider a system of steady electric currents in a finite space region of volume V , as shown in Fig. 8.5. Let $\mathbf{j}(\mathbf{r})$ be the current density in the position \mathbf{r} . Being that the regime is stationary, the current lines are closed. The current density is $\mathbf{j}(\mathbf{r}) = 0$ outside V .

As usual, the energy of the system is equal to the work spent to build it, starting from the state in which we define the energy to be zero. Let the null energy state be the state in which $\mathbf{j}(\mathbf{r}) = 0$ everywhere. Let us now have the current density increase to reach the final value. During this process, the magnetic field generated by the currents varies with time. The variation produces an electric field according to Faraday's law. In this phase of the process, in which the current density increases, generators are needed to work against the induced electric field, injecting energy into the system. We have already discussed similar situations in this chapter.

Let us calculate this work. We can consider having the current intensities vary slowly enough that the current lines are always closed, namely, in a formula, to have $\nabla \cdot \mathbf{j} = 0$ everywhere. The electric field induced by the current variations is given by Faraday's law:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}. \quad (8.14)$$

As we know, and as can be seen in Eq. (5.20), the work per unit time and unit volume done by the electric field \mathbf{E} on the current density \mathbf{j} is $\mathbf{E} \cdot \mathbf{j}$. Hence, the electric work of the generators in the infinitesimal time interval δt *against* the induced field is

$$\delta W = -\delta t \int_V \mathbf{E} \cdot \mathbf{j} dV = -\delta t \int_{\text{all space}} \mathbf{E} \cdot \mathbf{j} dV.$$

On the right-hand side, we could extend the integral to the entire space, not to V alone, because the integrand is zero outside V . We did that for reasons that will become clear immediately.

We now substitute \mathbf{j} using the expression

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j}, \quad (8.15)$$

obtaining

$$\delta W = -\frac{\delta t}{\mu_0} \int_{\text{all space}} (\nabla \times \mathbf{B}) \cdot \mathbf{E} dV.$$

We now use the vector identity

$$(\nabla \times \mathbf{B}) \cdot \mathbf{E} = -\nabla \cdot (\mathbf{E} \times \mathbf{B}) + (\nabla \times \mathbf{E}) \cdot \mathbf{B},$$

obtaining

$$\delta W = +\frac{\delta t}{\mu_0} \int_{\text{all space}} \nabla \cdot (\mathbf{E} \times \mathbf{B}) dV - \frac{\delta t}{\mu_0} \int_{\text{all space}} (\nabla \times \mathbf{E}) \cdot \mathbf{B} dV.$$

We see that the current system we are considering is spatially limited, namely that no current extends to infinity. Under this hypothesis, the first integral on the right-hand side, which is the integral of the divergence of a product of two fields, is zero. This statement can be proven with the same arguments we used in Sect. 3.5, which we shall not repeat here. In the second term, we use Eq. (8.14) and get

$$\delta W = +\frac{\delta t}{\mu_0} \int_{\text{all space}} \mathbf{B} \cdot \frac{\partial \mathbf{B}}{\partial t} dV = \frac{1}{\mu_0} \int_{\text{all space}} \mathbf{B} \cdot \frac{\partial \mathbf{B}}{\partial t} \delta t dV$$

But $\frac{\partial \mathbf{B}}{\partial t} \delta t$ is nothing more than the variation of \mathbf{B} in the time interval δt , which we call $\delta \mathbf{B}$. In addition, the work δW is just the variation in δt of the energy of the system, namely δU_m , and we can write

$$\delta W = \delta U_m = \frac{1}{\mu_0} \int_{\text{all space}} \mathbf{B} \cdot \delta \mathbf{B} dV \quad (8.16)$$

and also, obviously,

$$\delta U_m = \frac{1}{2\mu_0} \int_{\text{all space}} \delta B^2 dV, \quad (8.17)$$

where δB^2 is the variation of B^2 in the time interval δt .

We finally find the energy of the system by integrating on time from the initial instant in which the field is zero to the one in which the field has reached the final value, obtaining

$$U_m = \frac{1}{2\mu_0} \int_{\text{all space}} B^2 dV. \quad (8.18)$$

This important equation expresses the energy of the currents as the energy of the field created by those currents in the entire space. Note that field energy exists everywhere where $\mathbf{B} \neq 0$, namely also in regions in which there are no currents. The quantity $(B^2/2\mu_0)dV$ is the energy stored in the space volume dV . Hence, the energy is distributed in the space with the density

$$w_m = \frac{B^2}{2\mu_0}. \quad (8.19)$$

The different expressions for magnetic energy we have found in this chapter are equivalent to one another, as long as we are under time-independent conditions. They are no longer so when the quantities vary with time. Under these dynamic conditions, only Eq. (8.18) [and obviously Eq. (8.19)] holds, as we shall see in Sect. 10.3.

Consider, as an example, a cylindrical solenoid of circular section of radius r , length $l \gg r$ and n loops per unit length, carrying the steady current of intensity I . Neglecting the fringe effects, the magnetic field is zero outside the solenoid and uniform inside with magnitude $B = \mu_0 n I$.

Recalling that the inductance of the solenoid is $L = \mu_0 n^2 l \pi r^2$, the energy of the current generating the field is

$$U_m = \frac{1}{2} L I^2 = \frac{1}{2} \mu_0 n^2 I^2 l \pi r^2. \quad (8.20)$$

On the other hand, the energy stored in the field, neglecting the fringe effects, is simply the product of $B^2/2\mu_0$, which is uniform inside, times the volume of the solenoid, namely

$$U_m = \frac{B^2}{2\mu_0} l \pi r^2 = \frac{1}{2} \mu_0 n^2 I^2 l \pi r^2, \quad (8.21)$$

which is equal to Eq. (8.20).

QUESTION Q 8.1. Do the analogous calculation for a toroidal solenoid. □

QUESTION Q 8.2. The highest energy particle physics collider is the Large Hadron Collider (LHC) at CERN. The charged particles, protons or nuclei, circulate in two opposite directions in a ring 27 km long, thanks to a ladder of 1232 superconducting electromagnets producing a vertical magnetic field of intensity $B = 8.3$ T

generated with a current of intensity $I \cong 11.8$ kA. Each magnet is $l = 14.3$ m long, producing the field in a section we can assume to be $A = 180$ cm². What is the amount of energy stored in a magnet? How much in the full ring? How much is the inductance of a magnet? (answers: 7 MJ, 9 GJ, 0.1 H). \square

Summary

In this chapter, we have learned the following principal concepts:

1. The energy of a current.
2. The energy of a system of currents.
3. The energy of a magnetic dipole (a loop) in a magnetic field.
4. The energy and the energy density stored in a magnetic field.

Problems

- 8.1. The real resistors have non-zero inductance. As a consequence, when one connects a resistor to the poles of a battery, the current intensity does not reach the steady value immediately. The process takes some time, during which the battery spends work against the back emf (in addition to what is dissipated by the Joule effect). Where does this work go?
- 8.2. Inside a straight solenoid of small diameter, $l = 25$ cm long, carrying a steady current, the energy density is 2 J/m³. How many amper-turns (namely, number of loops times current intensity) are needed?
- 8.3. An electromagnet of inductance $L = 2$ H is connected in series to a constant emf generator and a switch. The magnet is excited with a current $I = 10$ A. Would you dare to open the switch under these conditions?
- 8.4. A small metal bar can slide along two parallel metal rails in a magnetic field perpendicular to the plane of the rails. The bar is kept moving at a constant speed. Consider two cases: **a** the circuit is closed on a resistor, **b** the circuit is closed on the same resistor in series with an inductor. In which case is the mechanical power we must spend larger? Why?
- 8.5. A coil with area $A = 1$ cm² is made of $N = 10$ overlapped loops and carries a steady current of intensity $I = 100$ nA. We introduce the coil into a solenoid producing the field $B = 0.8$ T. What is the total work needed if the operation is done with the magnetic moment of the coil in the same direction and sense as **B**? What is it if the senses are opposite? What are the agents of the works?
- 8.6. A loop of 1 cm² area carries a current of 1 A. It is located in a magnetic field **B** generated by a steady current of 100 A in a nearby circuit. **B** is directed

normally to the loop in the same sense as its magnetic moment. We measure the mutual inductance coefficient M between the two circuits. What is the value of \mathbf{B} ? And its energy?

- 8.7. In an experimental ultra-high-field NMR apparatus, the superconducting magnet produces a field of intensity $B = 17.6$ T with a current intensity $I = 200$ A. The stored energy is 5 MJ. How much is the volume of the field? What is the inductance of the magnet coil?

Chapter 9

Magnetic Properties of Matter

Abstract This chapter deals with the magnetic properties of matter. We consider three classes of material: diamagnetic, paramagnetic and ferromagnetic. Two new vector fields are introduced, needed to describe magnetism in matter: magnetization, which is the magnetic moment per unit volume, and the auxiliary \mathbf{H} -field. The latter is somewhat similar to the \mathbf{D} -field in electricity but its role is much more relevant. The magnetic phenomena in matter are due to the behavior of the atomic and molecular constituents. Even if the correct laws are those of quantum physics, we try to give an approximate classical description. We conclude with a discussion of ferromagnetic materials and their uses.

In this chapter, we study the principal magnetic properties of matter. We start with a few simple experimental observations, which distinguish three classes of materials, called diamagnetic, paramagnetic and ferromagnetic. To be precise, more classes exist, but we shall limit our discussion to these three principal ones. The ferromagnetic materials (for example, Fe, Co, Ni, Gd and several alloys) can be magnetized even in the absence of an applied magnetic field, namely they can be made permanent magnets. Both paramagnetic and diamagnetic materials magnetize under the action of an applied magnetic field (even if in a much weaker manner than the ferromagnetics) in the positive and negative directions of the field, respectively.

We introduce two new vector fields. The first one is “magnetization”, which is the magnetic moment per unit volume \mathbf{M} . We shall see how magnetization is produced by microscopic currents inside atoms and molecules, called equivalent magnetization currents. The situation is similar to that of the dielectrics, for which we established the relation between polarization, namely the electric dipole moment per unit volume, and the polarization charges.

We start in Sect. 9.2 with the simplest case of uniform \mathbf{M} , in which the microscopic current density is on the surfaces only, and continue in Sect. 9.3 with the case of non-uniform \mathbf{M} , in which the microscopic current density is also in the volume. A separation of the roles of the macroscopic and microscopic currents leads to the definition of the second vector field, \mathbf{H} . We study its properties and learn, in particular, that the “magnetic charge” observed since ancient times on the poles of the permanent magnets is not a real physical charge, but simply the

opposite of the divergence of the magnetization vector \mathbf{M} . This feature gives a clear interpretation of the broken magnet experiment described in Chap. 6.

The macroscopic magnetic properties of matter are due to phenomena at the molecular and atomic levels. As such, they are ruled by quantum mechanical laws, which cannot be discussed at the level of this course. We shall, however, try to give a description based on classical concepts in Sect. 9.6, warning the reader that it can only be approximated and therefore logically unsatisfactory. Still, it is useful for grasping the physics of the phenomena.

In Sects. 9.7, 9.8 and 9.10, we come back to ferromagnets, study the important phenomenon of the magnetic hysteresis and see how electromagnets produce strong magnetic fields. In Sect. 9.9, we give the microscopic interpretation of ferromagnetism.

Finally, in Sect. 9.10, we find the expressions of macroscopic energy, both in terms of the conduction currents and as stored in the field.

9.1 Elementary Observations

We start with a few simple observations, similar to those we did in Sect. 4.5 with the electric pendulum. We build a *magnetic pendulum*, attaching a sample of the material under study to a thin wire. We introduce the pendulum in the gap between the North and South poles of a powerful electromagnet. If the polar faces are plane and parallel, as in Fig. 9.1a, producing a uniform field in the location of the sample, we do not observe any effect, even if the field is very high. The effect is seen if one of the poles is shaped like the one in Fig. 9.2b. The magnetic field is not uniform now, being much stronger near the tip of the S pole. Up to now, the behavior is similar to that in the electric case, but there are important differences as well.

In the electric case, the pendulum is attracted toward higher electric field regions; in the magnetic case, we observe a mild attraction for some materials, called *paramagnetic*, and an even milder repulsion for others, called *diamagnetic*. A third class of materials exists, for which the attraction towards higher field regions is very intense, namely the *ferromagnetic materials*. From the analogy with the electric case, we can conclude that the magnetic field in any case magnetizes the sample, namely induces in the sample a magnetic dipole in the direction of the field.

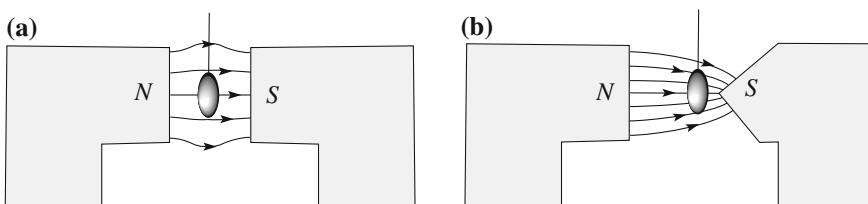
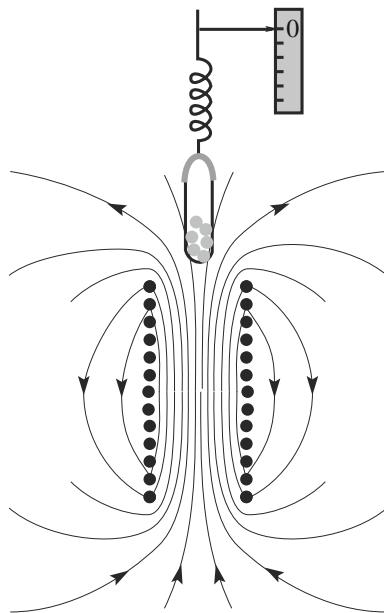


Fig. 9.1 Magnetic pendulum in **a** uniform, **b** non-uniform magnetic field

Fig. 9.2 Measuring the magnetic force on different samples



As opposed to the electric case, the induced magnetic moment has the same sense of the field for para- and ferro-magnetics, and the opposite for diamagnetics.

The arrangements in Fig. 9.1 are suitable only for semi-quantitative observations. Quantitative measurements can be done as shown schematically in Fig. 9.2. One employs a cylindrical solenoid producing a magnetic field of the requested intensity. The relatively simple geometry allows one to calculate the field and its gradient near the upper opening, where the sample to be measured is positioned. The sample is connected to a dynamometer to measure the force as a function of the field and of its gradient. To get an idea of the orders of magnitude, consider a field of 1 T intensity with a 15 T/m gradient, and samples of 1 g (whose weight, for comparison, is about 10^{-2} N). The intensities of the forces on a gram of diamagnetic substances (such as, for example, water, copper, bismuth, sodium chloride...) are on the order of 10^{-5} – 10^{-4} N, while on paramagnetic substances (such as sodium, aluminum, copper chloride,...), the forces are on the order of 10^{-4} – 10^{-2} N. Particularly strong is the force on liquid O₂ (at 90 K), which is 0.37 N under the above conditions, namely almost four times its weight.

As we shall see more clearly below, the molecules of a paramagnetic material have an intrinsic magnetic dipole moment. In the absence of an applied magnetic field, the elementary dipoles are randomly oriented. Consequently, the dipole moment of a piece of material containing a large number of molecules is zero. An applied magnetic field tends to give some orientation to the molecular dipoles, in contrast with thermic agitation. Everything is analogous to the polar dielectrics. In particular, paramagnetism depends considerably on temperature.

Contrastingly, the molecules of a diamagnetic material do not have a magnetic moment in the absence of an applied field, similarly to the non-polar dielectrics. An acting magnetic field, modifying the motion of the atomic electrons, induces a magnetic moment parallel to and in the opposite direction of the field. As we shall see, the opposite direction is a consequence of the Lenz law.

In a further analogy with dielectrics (positive direction apart), the induced polarization per unit volume at ordinary temperatures is usually larger by orientation (paramagnetic) than by deformation (diamagnetic).

The *ferromagnetic* materials are in a class of their own. In a non-uniform field, they are attracted toward higher field regions, much like the paramagnetics, but the force on ferromagnetic samples of the same mass is three to four orders of magnitude stronger, violently attracting the specimen inside the solenoid in the setup of Fig. 9.2. Under the conditions in the above-mentioned example ($B = 1$ T, $dB/dz = 15$ T/m), the attractive force on 1 g of Fe is about 2 N, namely about 200 times its weight, and on 1 g of magnetite (Fe_3O_4), 0.6 N.

Only ferromagnetic materials can be permanently magnetized. As we already mentioned, permanent magnets, such as magnetite, exist in nature, namely they were magnetized spontaneously. We can produce permanent magnets artificially by introducing a ferromagnetic sample into a magnetic field, inducing a magnetic moment. When we switch off the field, the magnetic moment decreases but does not disappear completely. We shall come back to ferromagnetism in Sects. 9.7–9.9.

9.2 Uniform Magnetization

We have mentioned that microscopic dipole moments, namely magnetically polarized molecules, are present in a magnetized medium. It is easy to understand that the magnetic field varies very rapidly in space over distances on the order of molecular dimensions. However, as in the case of dielectrics, we are interested here in the *macroscopic field*, namely in an average of the magnetic field over distances much larger compared to molecular dimensions (namely $\gg 0.1$ nm) and for times much longer than molecular times (namely $\gg 1$ fs). All the fields we shall consider are macroscopic fields.

The macroscopic magnetic field inside a material can be measured as follows. We use a beam of charged particles (think of electrons or protons) of energy high enough to be able to pass through the sample we are analyzing. We measure the direction of the beam before entering the material and at its exit, and consequently the deflection angle. The beam has been deflected by the Lorentz force $qv \times \mathbf{B}$ due to the field inside the material. Knowing the velocity and the charge of the particles of the beam, we can extract the average value \mathbf{B} .

We now define the (axial) vector *magnetization density*, or simply *magnetization*, \mathbf{M} as

$$\mathbf{M} = \frac{d\mu}{dV}, \quad (9.1)$$

where $d\mu$ is the magnetic moment of the volume dV (dV is extremely small from the macroscopic point of view, but still large enough to contain an enormous number of molecules). \mathbf{M} is the resultant magnetic moment per unit volume. In general, \mathbf{M} is a function of the coordinates and of time. However, we shall deal here only with time-independent conditions and we shall start with a uniform field.

Consider a cylindrical bar uniformly magnetized in the direction of the axis. On purpose, we can use a permanent magnet or magnetize an iron bar by introducing it into a solenoid carrying a steady current and producing a uniform field.

Let z be a coordinate axis parallel in the direction and sense of \mathbf{M} . Let us ideally cut a slice of thickness dz perpendicular to the axis, as shown in Fig. 9.3. Let us further divide the slice into square elements of infinitesimal area $d\Sigma$. Each of them has the magnetic moment $d\mu = M d\Sigma dz$.

Each of these elements creates a magnetic field in the surrounding space. This is equal to the field of a current loop having the shape of the lateral surface of the element and carrying a current with the right intensity to produce the same magnetic field (see Fig. 9.3). We call this an equivalent circuit. Its current intensity must be such that $dI d\Sigma = d\mu = M d\Sigma dz$, namely, simplifying, that $dI = M dz$. The equivalent circuit is a square ribbon of height dz . Properly speaking then, the current is a surface current, of surface intensity (measured in A/m)

$$k_m = \frac{dI}{dz} = M. \quad (9.2)$$

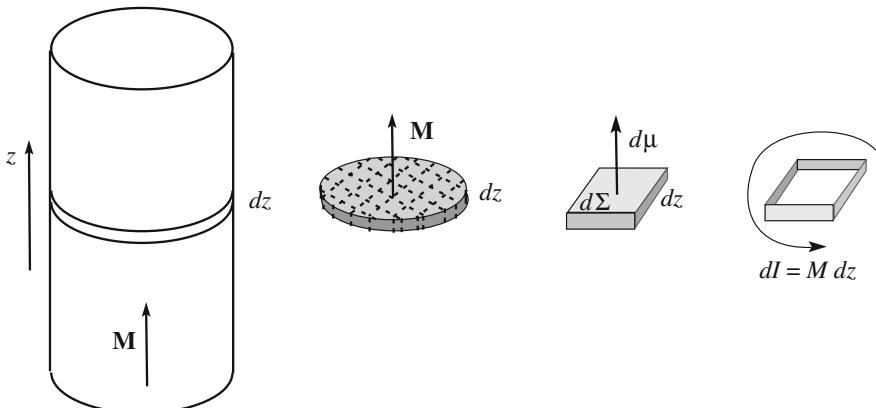


Fig. 9.3 A uniformly magnetized cylinder, a slice of the cylinder, an infinitesimal element of the slice and its equivalent circuit

If we substitute this into each element of the slice in Fig. 9.3, the equivalent loop we obtain will be like that shown in Fig. 9.4. We see that on each internal side, there are two equal and opposite currents, the effects of which cancel one another out. Note that this is the case because \mathbf{M} is uniform.

We conclude that the circuit equivalent of the magnetized slice is a circular ribbon equal to its lateral surface carrying the current $dI = M dz$, as shown in Fig. 9.5.

Similarly, the circuit equivalent to the complete magnetized bar is a cylinder of the shape of its lateral surface. The generic element of height dz of the lateral surface carries a current $dI = M dz$.

From this relation, we see, in particular, that the measurement units of magnetization are the ampere per meter (A/m).

We now notice that the equivalences we have established, between magnetized volumes and surface current densities on their lateral surface, holds only in the limit in which the equivalent current can be approximated with a magnetic dipole, namely far enough from the current itself.

However, we can also extend the equivalence inside the body. Let us continue to consider our cylindrical bar of uniform magnetization \mathbf{M} and the equivalent current of surface density $k_m = M$. We know that the magnetic fields of the two are equal outside. Let us consider, in both cases, a closed surface S enclosing one face of the cylinder, as shown in Fig. 9.6. Let S_e be its part outside the cylinder and S_i its part inside. The flux of \mathbf{B} outgoing from S is zero in both cases. The flux of \mathbf{B} through S_e being equal in both cases, we must conclude that the flux through S_i is equal in both cases as well. But the flux of \mathbf{B} through S_i is the average value of \mathbf{B} inside times the area of S_i . Consequently, the averages of \mathbf{B} are also equal inside the cylinders.

Fig. 9.4 The infinitesimal circuits equivalent to the elements of the slice

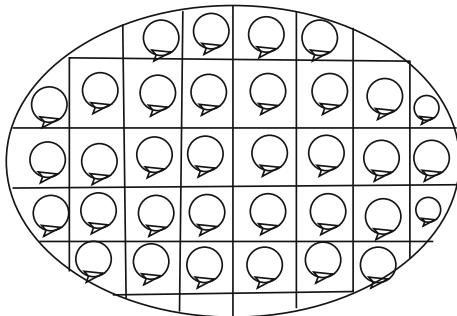
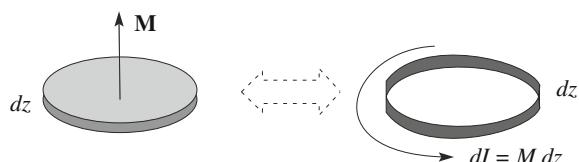


Fig. 9.5 The slice and its equivalent circuit



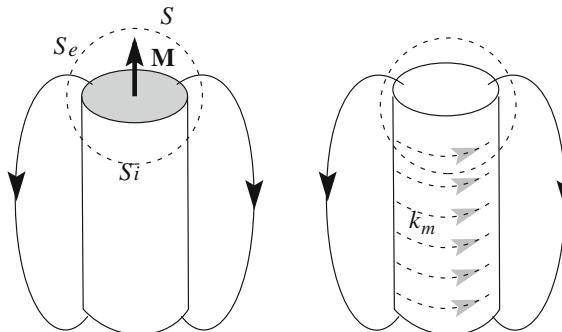


Fig. 9.6 Comparing the magnetic field of the magnetized bar and the equivalent current

Consider a cylindrical permanent magnet. Its magnetic field is equal to the field of a solenoid (namely a surface current) of the same geometry, with a current per unit length (namely the product of the current intensity times the number of turns per unit length) equal to the magnetization density of the permanent magnet.

9.3 Non-uniform Magnetization

Consider now a case in which the magnetization \mathbf{M} does not depend on time but depends on position. We shall see that the equivalent currents, which are only the surfaces in the uniform case, are also present in the bulk in the non-uniform case.

Let us start by dividing the volume V of the body into infinitesimal cubic elements $dV = dx dy dz$. Let us consider two such contiguous elements along the y -axis, at y and $y + dy$, respectively, as represented in Fig. 9.7a, and the z -components of their magnetization, namely $M_z(y)$ and $M_z(y + dy)$. As we already know,

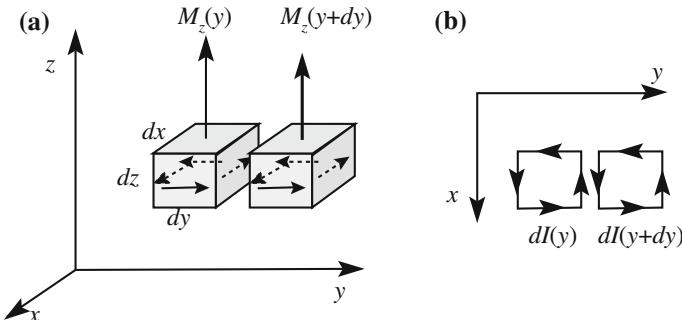


Fig. 9.7 a Two contiguous magnetized elements along the y -axis, the z -components of their magnetization; b the equivalent circuits

the two blocks are equivalent to two current ribbons having the shape of their lateral surfaces (of height dz) carrying the currents $dI(y) = M_z(y)dz$ and $dI(y + dy) = M_z(y + dy)dz$.

As shown in Fig. 9.7b, the two opposite currents running on the common face of the two circuits, as opposed to the uniform case, now have different intensities and do not cancel one another out. A net current in the x -direction remains (to which we give a positive sign if it has the positive x direction). This current intensity is infinitesimal of the second order and is given by

$$d^2I = [M_z(y + dy) - M_z(y)]dz = \frac{\partial M_z}{\partial y}dydz.$$

We must consider that, in the y -direction, there is a contribution analogous to the one just considered every dy , and in the z -direction every dz . In other words, there is a contribution d^2I every infinitesimal square $dy dz$. Hence, we can speak of a current density whose x -component is

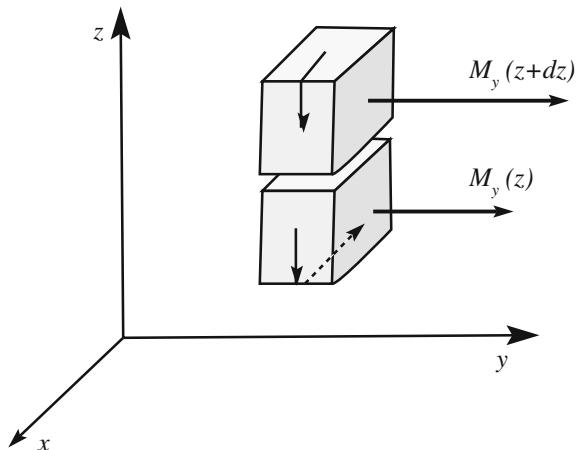
$$j_x = \frac{d^2I}{dydz} = \frac{\partial M_z}{\partial y}.$$

There is, however, another contribution to j_x . It comes from the currents equivalent to the y -components of the magnetization on the common faces of two contiguous blocks in the z -direction, as shown in Fig. 9.8.

Looking at the figure, we see (paying attention to the sign) that the contribution of the common surface of two blocks to the current in the positive x -direction is

$$d^2I = [M_y(z) - M_y(z + dz)]dy = -\frac{\partial M_y}{\partial z}dzdy.$$

Fig. 9.8 Two contiguous magnetized elements along the z -axis, the y -components of their magnetization



This current corresponds to the square surface of area $dzdy$. Consequently, its contribution to the current in the x -direction is

$$j_x = \frac{d^2I}{dydz} = -\frac{\partial M_y}{\partial z}.$$

In conclusion, adding the two contributions, the x -component of the *magnetization current density* j_m is

$$j_{mx} = \frac{\partial M_z}{\partial y} - \frac{\partial M_y}{\partial z},$$

which we immediately recognize as the x -component of $\nabla \times \mathbf{M}$. Obviously, similar results hold for the components on the y and z axes, and we can say that the magnetization current density \mathbf{j}_m corresponding to the magnetization \mathbf{M} is

$$\mathbf{j}_m = \nabla \times \mathbf{M}. \quad (9.3)$$

Let us now look at how the previously considered case of uniform magnetization is a special case of what we have just discussed. Let us consider again the cylindrical bar uniformly magnetized in the direction of its axis, which we take to be the z -axis. The curl of \mathbf{M} is zero inside, where \mathbf{M} is uniform, and \mathbf{j}_m is zero as well. Crossing the surfaces, \mathbf{M} , or better yet, its only non-zero component M_z , varies abruptly from the value it has inside to zero. The non-zero partial derivative crossing the two faces is $\partial M_z / \partial z$, which does not contribute to the curl. Consequently, the current density is zero on the faces. The curl of the magnetization is different from zero only on the lateral surface. In Sect. 9.2, we established that the current equivalent to the magnetization in this case is just a surface current on the lateral surface of the cylinder of density $k_m = M$. We can now see that the two points of view are equivalent thinking that every physical surface has a thickness that may be small, but is never null. Let δ be the thickness of the current-carrying surface. Then, the relation between volume and surface current densities is $j_m = k_m / \delta$.

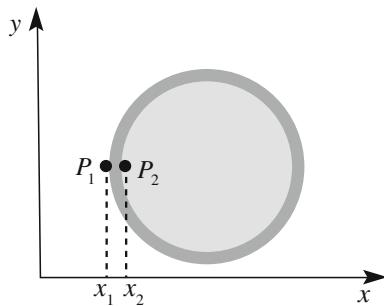
Figure 9.9 represents a section of the bar and two points, one, P_2 , immediately inside, the other, P_1 , immediately outside. The two points have the same y and coordinates x_1 and x_2 separated by the distance δ . At both points, \mathbf{j}_m and the corresponding \mathbf{k}_m are directed along the y -axis. As $k_m = M_z$, the arguments of Sect. 9.2 lead us to expect that $j_{my} = M_z / \delta$.

On the other hand, Eq. (9.3), which we want to show to be equivalent, says that

$$j_{my} = (\nabla \times \mathbf{M})_y = -\frac{\partial M_z}{\partial x} = \frac{M_z(x_1) - M_z(x_2)}{\delta} = \frac{M}{\delta} = \frac{k_m}{\delta},$$

which coincides with the above expression.

Fig. 9.9 A cross-section of the magnetized bar and two points on the sides of its lateral surface



9.4 Magnetic Field Equations in Matter

In this section, we write down all the equations describing the magnetic field in matter under time-independent conditions, namely when the charge density and the current density are constant in time. In the next chapter, we shall see how the equations are generalized under dynamic conditions. We already know the two differential equations of the magnetic field, namely

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} \quad (9.4)$$

and

$$\nabla \cdot \mathbf{B} = 0. \quad (9.5)$$

These equations hold both in a vacuum and in the presence of matter. In the latter case, however, beyond the conduction currents, which are macroscopic, microscopic currents also exist, in the volume and on the surfaces, which are not under our direct control. It is consequently convenient to separate the current density \mathbf{j} into two parts. One part, \mathbf{j}_c , is the conduction current, which is under our control; the other is the magnetization current \mathbf{j}_m . Namely, we write

$$\mathbf{j} = \mathbf{j}_c + \mathbf{j}_m = \mathbf{j}_c + \nabla \times \mathbf{M} \quad (9.6)$$

Note that the electric polarization charges that may be present are steady, because we are in a time-independent situation and consequently do not contribute to the microscopic current. This would not be so under dynamic conditions, as we shall see in the next chapter. We now insert this expression into Eq. (9.4) and find

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j}_c + \mu_0 \nabla \times \mathbf{M}$$

and, moving \mathbf{M} to the left-hand side and dividing by μ_0 , we obtain

$$\nabla \times (\mathbf{B}/\mu_0 - \mathbf{M}) = \mathbf{j}_c. \quad (9.7)$$

It is then useful to introduce the auxiliary field \mathbf{H} defined as

$$\mathbf{H} = \mathbf{B}/\mu_0 - \mathbf{M}. \quad (9.8)$$

The measurements units of \mathbf{H} are the same as of \mathbf{M} , namely the ampere per meter (A/m). The two magnetostatic field equations become

$$\nabla \times \mathbf{H} = \mathbf{j}_C \quad (9.9)$$

and

$$\nabla \cdot \mathbf{B} = 0. \quad (9.10)$$

In this form, which is obviously equivalent to the previous one, the right-hand sides of the equations are known. Obviously, the complications due to matter still exist; they have been moved to the left-hand side of Eq. (9.9) and hidden in the definition of \mathbf{H} . Note that the \mathbf{H} field depends on the macroscopic currents only. The situation is similar to that of electrostatics, where we introduced the field \mathbf{D} , the sources of which are the free charges alone. However, while \mathbf{D} is not very helpful in practice, \mathbf{H} is quite useful. Indeed, in electrostatic situations, we usually control the potentials of the conductors, rather than their free charges. The electric field \mathbf{E} is, in any case, the opposite of the gradient of the potential, both in the presence and absence of dielectrics. Consequently, we directly control \mathbf{E} , not \mathbf{D} . For example, the electric field between the plates of a parallel plate capacitor is the potential difference divided by the distance between the plates, both with a vacuum and with a dielectric. Contrastingly, in the magnetic case, we control the conduction currents, hence \mathbf{H} rather than \mathbf{B} .

Equation (9.9) is in a differential form. It can be written in an integral form as well, stating that the circulation of \mathbf{H} about any oriented curve Γ is equal to the macroscopic current I_C linked to that curve (with the usual convention for the sign), namely

$$\oint_{\Gamma} \mathbf{H} \cdot d\mathbf{s} = I_C \quad (9.11)$$

The equations we have found are useless as long as we do not establish a relation between \mathbf{B} (the cause) and the magnetization \mathbf{M} (the effect). This relation is, in principle, a difficult one, because it must describe the behavior of the microscopic matter constituents under the action of an external magnetic field. Consequently, the relation can only be an approximation. Fortunately, however, a linear approximation is extremely good for the dia- and para-magnetic materials. Unfortunately, for historical reasons, this relation is always written in terms of \mathbf{H} rather than of \mathbf{B} , namely as

$$\mathbf{M} = \chi_m \mathbf{H}. \quad (9.12)$$

The dimensionless constant χ_m is called the *magnetic susceptibility*. We repeat that this relation holds for dia- and para-magnetics, but not, in general, for ferromagnetic materials. This is a consequence of the fact that the magnetic susceptibility of the former materials is always very small in absolute value, namely $|\chi_m| \ll 1$. Recalling that \mathbf{M} has the direction of \mathbf{B} and its sense in paramagnetics, the opposite of its sense in diamagnetics, we see that $\chi_m > 0$ for the former, $\chi_m < 0$ for the latter ones.

We can now write Eq. (9.8) as

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M}) = \mu_0(1 + \chi_m)\mathbf{H} = \kappa\mu_0\mathbf{H} = \mu\mathbf{H}, \quad (9.13)$$

where we have defined two constants characteristic of the medium, namely

$$\kappa = 1 + \chi_m, \quad (9.14)$$

which is dimensionless and is called the *magnetic permeability* of the medium *relative* to the vacuum, and

$$\mu = \kappa\mu_0 = \mu_0(1 + \chi_m), \quad (9.15)$$

called the *absolute magnetic permeability*. The measurement units of μ are the same as of μ_0 , namely the newton per ampere square, N A^{-2} .

The linear relation Eq. (9.12) between M and H impels a linear relation between M and B as well, namely

$$\mathbf{M} = \frac{\chi_m}{\mu_0(1 + \chi_m)} \mathbf{B}. \quad (9.16)$$

As a matter of fact, this relation, rather than Eq. (9.12), should have been the starting point of the above arguments, because the fundamental field is \mathbf{B} and \mathbf{H} is an auxiliary field. Indeed, as we have already mentioned, a high energy charged particle beam crossing a magnetized material senses \mathbf{B} through the Lorentz force, and not \mathbf{H} . It would have been more logical to start from the definition

$$\mathbf{M} = \frac{\chi'_m}{\mu_0} \mathbf{B}. \quad (9.17)$$

The reason for why it is done as we did it is historical. Indeed, up to the middle of the previous century, \mathbf{H} was believed to be a fundamental field. In practice, however, in the non-ferromagnetic materials we are considering, the complication is immaterial. Indeed, being that $|\chi_m| \ll 1$, it follows that

Table 9.1 Magnetic susceptibility of diamagnetic substances in 10^{-5} units at STP

Substance	H_2O	Bi	C (diamond)	Cu	Ag	NaCl
χ_m	-0.9	-17	-2.2	-1.0	-2.6	-1.4

Table 9.2 Magnetic susceptibility of paramagnetic substances in 10^{-5} units at STP (liquid O_2 at 73 K)

Substance	Al	Cs	Na	CuCl_2	Liquid O_2	O_2	Air
χ_m	+2.2	+5.1	+0.72	+34	+390	+0.19	+0.036

$$\chi'_m = \frac{\chi_m}{1 + \chi_m} \cong \chi_m. \quad (9.18)$$

In these cases, the two constants are almost identical. This is not true for the ferromagnets.

Tables 9.1 and 9.2 report the measured magnetic susceptibilities for a number of diamagnetic and paramagnetic substances, respectively. When mentally comparing the values, keep in mind that what is proportional to the inducing magnetic field through χ_m is the induced magnetic moment *per unit volume* and that the number of molecules in a cubic meter at STP (standard temperature and pressure) is typically three orders of magnitude smaller in gases (such as oxygen and air) than in condensed media.

9.5 B and H Fields in Matter

Let us consider once more a cylindrical bar uniformly magnetized in the direction of its axis. This means that every volume element dV of the medium has a magnetic dipole $d\mu = M dV$. The magnetization \mathbf{M} is uniform inside the bar and equal to zero outside.

Let us discuss the three fields \mathbf{B} , \mathbf{M} and \mathbf{H} . Suppose we are not dealing with a permanent magnet. We insert the cylinder into the uniform field \mathbf{B} of a solenoid. The induced \mathbf{M} is uniform and proportional to the field \mathbf{B} , in its own sense for the paramagnetic materials, in the opposite sense for the diamagnetics. In both cases, the magnetization is quite small, $\mu_0 M \ll B$. Let us consider the different types of material.

Paramagnetic materials.

Figure 9.10a represents the empty solenoid, producing the field \mathbf{B} in a vacuum. When we introduce the sample with $\chi_m > 0$, the induced magnetic moment is in the direction and sense of \mathbf{B} . From this point of view, the behavior is similar to that of dielectrics in which \mathbf{P} has the direction and sense of \mathbf{E} . The difference is that, while in a dielectric, the magnitude of \mathbf{E} is reduced inside the medium, in the paramagnetics, \mathbf{B} is larger in the presence than in the absence of the material. Indeed, the

magnetization is equivalent to a surface current, which has the same direction as the macroscopic current in the solenoid. Consequently, the field lines are denser inside the bar than outside it. The bar looks to be more “permeable” to the field compared to the vacuum, hence the name “magnetic permeability”. In practice, the effect is always small and we have exaggerated it in the figure to make it more visible.

There is not too much to be said on \mathbf{H} , because this field is proportional to \mathbf{B} [see Eq. (9.13)].

Diamagnetic materials.

In this case, the sense of the induced magnetic field is opposed to that of the inducing field \mathbf{B} , because $\chi_m < 0$. From this point of view, the behavior is different compared to the dielectrics. However, as in the dielectrics, \mathbf{B} is now weaker when the medium is present. Indeed, inside the medium, \mathbf{B} is the result of the macroscopic current and the equivalent magnetic current on the surface of the bar, which run in opposite directions. The field lines “prefer” running in a vacuum as much as possible. The medium is less “permeable” to them than the vacuum. In this case as well, we have exaggerated the effect in Fig. 9.11.

The effects are once again small, even smaller than in paramagnetics, and again, a discussion of \mathbf{H} is not relevant.

Fig. 9.10 **a** A solenoid creating a uniform magnetic field in a vacuum; **b** same with inserted paramagnetic material

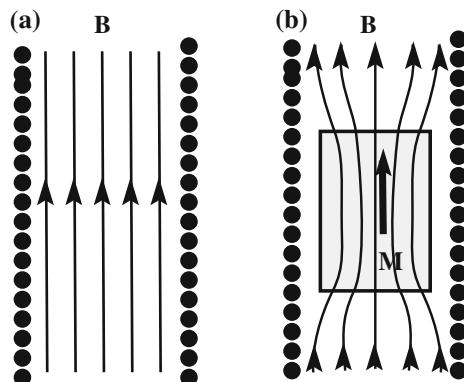
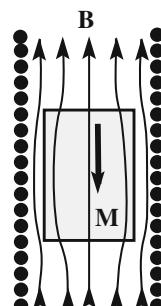


Fig. 9.11 A diamagnetic bar in a uniform field



Ferromagnetic materials.

A ferromagnetic bar behaves like a paramagnetic one, as far as the direction of the induced sense of \mathbf{M} relative to the inducing field \mathbf{B} is concerned. However, this is the only analogy between the two cases.

Firstly, for the same \mathbf{B} , the magnitudes of \mathbf{M} are much larger but also, even more importantly, if we switch off \mathbf{B} by decreasing the current in the solenoid to zero, we observe that \mathbf{M} decreases somewhat, but reaches a value, which is different from zero and possibly quite large (see Sect. 9.7). This is the already-mentioned phenomenon of permanent magnetization.

We have already seen that the magnetic field \mathbf{B} generated by our cylindrical magnetized bar is equal to the field of a solenoid of the same shape, carrying the surface current density $k_m = nI$ (where n is the number of turns per unit length and I the current intensity) equal to the bar magnetization M .

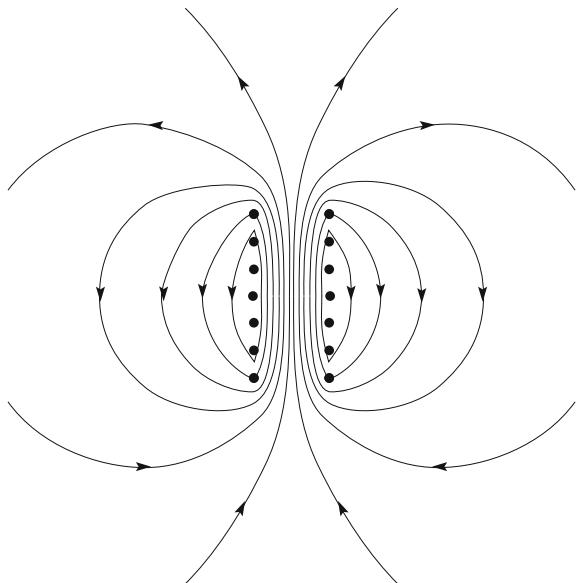
Figure 9.12 shows the \mathbf{B} field of our solenoid. It has the same shape as that of the bar in Fig. 9.14a. Note that \mathbf{B} is discontinuous crossing the lateral surface. Indeed, we know that, through a current sheet of density k , the normal component of \mathbf{B} is continuous, while the tangential one, say B_t , has the discontinuity

$$\Delta B_t = \mu_0 k. \quad (9.19)$$

Consider now the \mathbf{H} field. In this case, the simple proportionality relations Eq. (9.13) do not hold. However, \mathbf{M} is now given. It is uniform inside and zero outside the bar. The field \mathbf{H} is given by

$$\mathbf{H} = \mathbf{B}/\mu_0 - \mathbf{M}. \quad (9.20)$$

Fig. 9.12 The \mathbf{B} field of a cylindrical solenoid



In the case of the solenoid, being that it is a vacuum, \mathbf{H} is proportional to \mathbf{B} everywhere, namely it is $\mathbf{H} = \mathbf{B}/\mu_0$. \mathbf{H} is not very useful under these conditions.

Let us consider the cylindrical magnet. Outside the relation, $\mathbf{H} = \mathbf{B}/\mu_0$ is still true, and consequently, \mathbf{H} has the same shape as \mathbf{B} . Contrastingly, inside the magnet, we must also take into account the contribution of \mathbf{M} , which may be extremely important. Figure 9.13 shows an example of the two vector terms on the right-hand side of Eq. (9.20) and of the resulting left-hand side. Note, in particular, that the direction of \mathbf{H} may be very different from that of \mathbf{B} , as in this example.

Figure 9.14a, b represent the \mathbf{B} and \mathbf{H} fields of the cylindrical magnet, respectively. Outside, the shapes of the two fields are equal, while inside, they are completely different, so much so as to have opposite directions, on average. The field \mathbf{H} is discontinuous on the faces of the cylinder, which are the poles of the magnet. The \mathbf{H} field lines exit from the North pole and enter into the South pole. This behavior is very similar to that of the electrostatic field, whose lines exit the

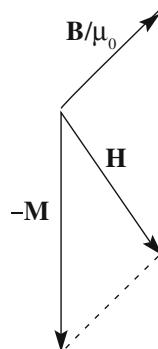


Fig. 9.13 The vectors \mathbf{B} , \mathbf{M} and \mathbf{H} inside a ferromagnet

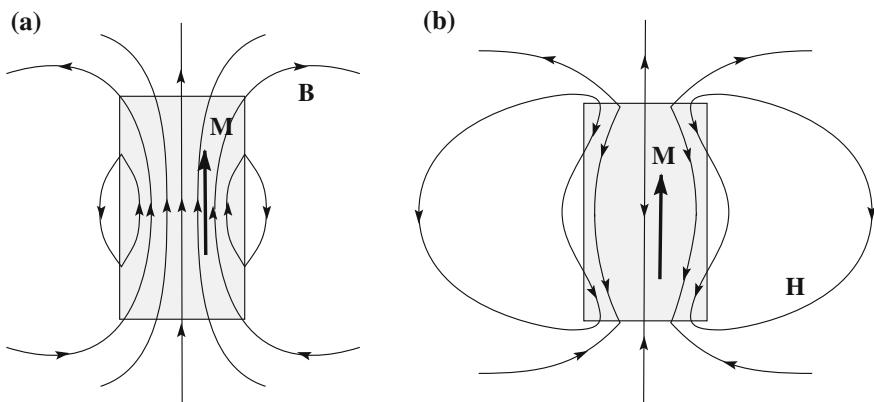


Fig. 9.14 Fields of a permanent magnet **a** \mathbf{B} , **b** \mathbf{H}

positive and enter into the negative charges (think of a positive charge density on the N pole and an opposite negative one on the S pole surfaces).

The analogy is very close, but purely formal. To see that, let us take the divergence of both sides of Eq. (9.20), recalling that $\nabla \cdot \mathbf{B} = 0$. We have

$$\nabla \cdot \mathbf{H} = -\nabla \cdot \mathbf{M}.$$

If we now define a *magnetic charge density*

$$\rho_m \equiv -\nabla \cdot \mathbf{M}, \quad (9.21)$$

the equation becomes

$$\nabla \cdot \mathbf{H} = \rho_m, \quad (9.22)$$

which is identical, a constant apart, to the electrostatic field equation

$$\nabla \cdot \mathbf{E} = \rho/\epsilon_0$$

As for the curl, in the case we are considering of a permanent magnet in the absence of conduction currents, Eq. (9.9) becomes

$$\nabla \times \mathbf{H} = 0, \quad (9.23)$$

namely, the \mathbf{H} field is conservative, as is the electrostatic field. These properties of the \mathbf{H} field of permanent magnets historically led physicists to assign an importance to the \mathbf{H} field much larger than it really deserves. Clearly, Eq. (9.22) is a purely formal definition of a *magnetic charge*. No physical magnetic charge exists. The magnetic charge density is just the opposite of the divergence of magnetization. Looking back at the example of the uniformly-magnetized bar, it is clear that $\nabla \cdot \mathbf{M} = \partial M_z / \partial z$ is different from zero only on the polar faces of the cylinder, where M_z changes suddenly from its internal value to zero. Consequently, the “magnetic charges” are concentrated at the poles. We can verify that easily by dropping the bar into a pot of iron filings. We see them attaching to the poles. Obviously, the values of $\partial M_z / \partial z$ are equal and opposite on the two faces, and consequently, the “magnetic charges” are equal and opposite as well.

The same argument also explains the broken magnet experiment, which we mentioned in Sect. 6.1. If we cut a magnetized stick in the middle, we produce two new faces, one for each of the parts. On them, \mathbf{M} has sudden variations and a sizeable divergence. We have created two new “magnetic charges”, one on each side equal and opposite in sign.

In closing this section, we observe that the discontinuities of the \mathbf{H} field are similar to those of the \mathbf{E} field, because both fields have the same divergence and the same curl. Crossing a magnetic charge surface, the tangent component of \mathbf{H} is continuous, while its normal component is discontinuous.

9.6 Dia- and Para-magnetism. Microscopic Interpretation

The magnetic properties of matter have their physical origin in its microscopic constituents, namely molecules, atoms, electrons and nuclei. At this level, physics follows quantum laws, of which the classical ones are a good approximation only at scales substantially larger than the atomic ones. Consequently, only quantum physics can properly explain paramagnetism, diamagnetism and ferromagnetism. Historically, the development of quantum mechanics took a few decennia and the work of several scientists. One of the major contributors was Niels Bohr (Denmark, 1885–1962), who proposed, in 1913, an atomic model, in which the atom is similar to a planetary system, with the nucleus at the center and the electrons moving about it in closed orbits. The centripetal force is the electric attraction of the positive nucleus for the negative electron. As opposed to the solar system, the electron orbits are in different planes, in all directions. This is the usual, but not physically correct, image of the atom. The model was, and still is, capable of explaining a number of observed phenomena, but gives incorrect predictions as well. The main reason for its failure is that, in quantum mechanics, the trajectory of a particle can be defined only within a well-defined uncertainty, which is that much larger the larger the momentum of the particle. The Bohr model was a historically important step forward in the construction of quantum mechanics, similar to the scaffolding used to build a cathedral. Once the cathedral is built, the scaffolding is removed.

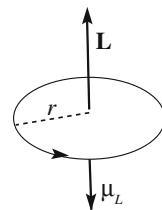
We shall, however, use the Bohr model to give a semi-quantitative microscopic interpretation of the magnetic phenomena. We shall clearly state when the predictions of the model are somewhat justified by their agreement with the quantum predictions and when this is not the case.

Two important quantities, closely linked to one another in both quantum and classical physics, are the angular momentum and the magnetic dipole moment. We can distinguish two types of angular momenta of an atomic electron: the *orbital angular momentum* and, as it is called, the *spin*. The former, which we call \mathbf{L} , has a classical analogy; the latter does not. If the electron was a small charged sphere quickly spinning about its axis, it would classically have an angular momentum and a magnetic moment due to that spinning. However, electrons are point-like. The spin, which we shall call \mathbf{S} , is an intrinsic, purely quantum property of the particles.

Using the Bohr model, we now consider an electron moving along a circular orbit of radius r with velocity \mathbf{v} (which is much smaller than the speed of light). The orbital angular momentum of the electron is the angular momentum resulting from this motion. It is normal to the plane of the orbit, in the direction that sees the electron moving counterclockwise, and has the magnitude $L = m_e v r$, where m_e is the mass of the electron (see Fig. 9.15).

Moving along its orbit, the electron produces a current, which is running along the orbit itself. Its intensity is the electron charge q_e times the number of times the electron crosses any given point in a second, which is $v/(2\pi r)$. Hence, the current

Fig. 9.15 Orbital angular momentum and dipole magnetic moment of an atomic electron



intensity of one electron is $I = q_e v / (2\pi r)$. Being that the electron charge is negative, the direction of the current is opposed to the velocity of the electron. The orbit is a small current loop, namely a magnetic dipole with a magnetic moment parallel and opposite to the angular momentum, with magnitude

$$\mu = \frac{q_e v}{2\pi r} \pi r^2 = \frac{q_e}{2m_e} m_e v r = \frac{q_e}{2m_e} L.$$

We conclude that the orbital angular momentum and the magnetic moment are two parallel and opposite (axial) vectors, proportional to one another according to the relation

$$\boldsymbol{\mu}_L = -\frac{q_e}{2m_e} \mathbf{L}. \quad (9.24)$$

This result is very important. Note that the proportionality constant between the orbital angular momentum and the magnetic moment $-q_e/(2m_e)$ depends only on the properties of the electron. As such, it is universal, independent, in particular, of the shape of the orbit. The latter might not even be completely defined. As a matter of fact, our result, even if obtained with arguments of classical physics, is a posteriori justified by the fact that it is also foreseen by quantum mechanics.

As anticipated, this is not the case for the intrinsic angular momentum, namely the spin, of the electron, which is a purely quantum phenomenon. A magnetic moment, which we call $\boldsymbol{\mu}_S$, is associated with the spin as well. The two (axial) vectors are parallel and in opposite direction. Their magnitudes are proportional, but the proportionality factor is twice as large as for the orbital case, namely it is

$$\boldsymbol{\mu}_S = -\frac{q_e}{m_e} \mathbf{S}. \quad (9.25)$$

Let us now consider an atom with a number of electrons. Each of them has its orbital and spin angular momenta and the associated magnetic moments. Their directions are different. The total angular momentum of the electron system, which we call \mathbf{J} , and its total magnetic moment, $\boldsymbol{\mu}_J$, are the vector sums of the contributions of all the electrons. The vector sum rules obey quantum, not classical, mechanics. The result, which has no classical explanation, is that \mathbf{J} and $\boldsymbol{\mu}_J$ are still parallel and opposed, linked by the relation

$$\mu_J = -g \frac{q_e}{m_e} \mathbf{J}, \quad (9.26)$$

where the dimensionless constant g is called the Landé factor, after Alfred Landé (Germany, 1888–1976), who first discussed it. Each atomic species has its Landé factor, whose calculation is a task of quantum mechanics. The Landé factors are on the order of the unit. For completeness, it should be mentioned that the atomic nucleus has an angular momentum and a magnetic moment too, but these do not contribute to the phenomena we are discussing.

We are now ready to discuss diamagnetic and paramagnetic phenomena.

Diamagnetism

In the atoms and molecules of diamagnetic substances, the orbital and spin electron angular momenta add up to zero. This fact is quite ordinary, because electrons tend to form couples with total angular momentum equal to zero. In these cases, the total angular momentum and the total magnetic moment are zero, as long as the system is not perturbed.

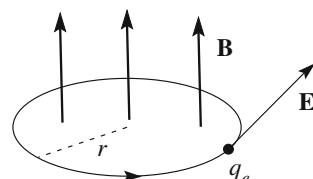
When the atom is in an external magnetic field, its state is different from that of the unperturbed one. To understand the issue, let us now assume that we have the field \mathbf{B} , gradually growing from zero to its final value. Our argument will, again, be classical. Let us consider an orbit, which is also our circuit, normal to the field, as in Fig. 9.16, of radius r . As the magnetic flux linked to the orbit varies, an electric field \mathbf{E} is induced (Faraday's law) at all the points of the orbit (in particular). The circulation of \mathbf{E} about the orbit is equal to the opposite of the rate of change of the linked flux of \mathbf{B} . We can certainly consider \mathbf{B} to be uniform on the extremely small size of the orbit and write that $E2\pi r = -d(B\pi r^2)/dt$. Considering r to be constant, the induced electric field is then

$$E = -\frac{r}{2} \frac{dB}{dt}$$

The force exerted by the field on the electron ($-q_e E$) is tangential to the orbit. Its moment about the center is $-q_e Er$, is directed normal to the orbit, and gives the rate of change of the electron angular momentum

$$\frac{dL}{dt} = -q_e Er = \frac{q_e r^2}{2} \frac{dB}{dt}$$

Fig. 9.16 An electron “orbit” normal to \mathbf{B} , the linked magnetic flux and the induced electric field acting on the electron



The total change in the electron orbital angular momentum, ΔL , when the magnetic field has reached the final value B is immediately obtained by integration of this equation, obtaining $\Delta L = (q_e r^2 / 2)B$. Correspondingly, the electron has gained, relative to the unperturbed state in the absence of a magnetic field, an extra corresponding magnetic moment $\Delta \mu$. As we are considering the orbital momenta, the proportionality constant is $-q_e/(2m_e)$, and we have

$$\Delta \mu = -\frac{q_e}{2m_e} \Delta L = -\frac{q_e^2 r^2}{4m_e} \mathbf{B}, \quad (9.27)$$

Notice that the minus sign is substantially a consequence of the Lenz law. Namely, the extra magnetic moment $\Delta \mu$, resulting from the change in the orbital motion, tends to oppose the cause that generated the effect, which is the growth of \mathbf{B} , and consequently, $\Delta \mu$ is opposite to \mathbf{B} . Again, this result is validated by being in accord with quantum mechanics. This substantially explains diamagnetism; namely, the induced magnetic moment is proportional to the magnetic field and in the same direction and opposite sense.

Consider now orbits that are not normal to the field. The argument is the same if we take, in the place of \mathbf{B} , its component perpendicular to the orbit. Let us consider a frame with its origin in the center and the z -axis in the direction of \mathbf{B} . In Eq. (9.27), we must write, in place of the square of the orbit radius $r^2 = x^2 + y^2 + z^2$, the square of its projection normal to \mathbf{B} , namely $r'^2 = x^2 + y^2$. In calculating the macroscopic effects, we must consider an average on all the orbit orientations. Clearly, the directions being randomly distributed, we have $\langle x^2 \rangle = \langle y^2 \rangle = \langle z^2 \rangle$, and consequently, $\langle r'^2 \rangle = \langle x^2 \rangle + \langle y^2 \rangle = 2\langle r^2 \rangle / 3$. We then write in place of Eq. (9.27)

$$\Delta \mu = -\frac{q_e^2 \langle r^2 \rangle}{6m_e} \mathbf{B}, \quad (9.28)$$

where $\langle r^2 \rangle$ is the mean value of the orbit square radius. The induced magnetic moment per unit volume is immediately obtained by multiplying this average value times the number of molecules per unit volume n_p , namely

$$\mathbf{M} = -\frac{q_e^2 n_p \langle r^2 \rangle}{6m_e} \mathbf{B}. \quad (9.29)$$

We have found that \mathbf{M} is proportional to and opposite of \mathbf{B} . Diamagnetism is a property of substances whose atoms have all their electronic shells filled with electrons, because all the electrons are then paired. Such are He and the noble gases. The singly ionized monovalent metals (Li, Na, Cs, etc.) are diamagnetic for the same reasons. In the vast majority of cases, the susceptibility is quite small, typically on the order of 10^{-5} , as one can see in the examples in Table 9.1.

Paramagnetism

The pairing of atomic electrons in couples of equal and opposite angular momenta is clearly not possible if the number of electrons is odd. More cases exist in which some of the electrons are un-paired. In all these cases, the atom (or the molecule) has a non-zero intrinsic magnetic moment, namely a moment existing even in the absence of an acting field. These are the paramagnetic substances.

However, in the absence of an external field in any (macroscopically) infinitesimal volume (very small but still containing a large number of molecules), the net magnetic moment is zero. The reason is the same as that for the polar dielectrics. In their thermic statistical equilibrium, the molecular magnetic moments μ_J are casually oriented, with the same probability in all directions. In the presence of a field \mathbf{B} , the magnetic moments tend to orient in the direction and sense of \mathbf{B} (in the same sense, because both μ_J and \mathbf{B} are constant). The thermic motion opposes this ordering process. This situation is identical to that of the polar dielectrics. The probability of finding a molecular dipole μ at the angle θ with the field \mathbf{B} is given by the Boltzmann factor (see Volume 2, Chap. 5) as a function of the temperature and the energy of the dipole at that angle, namely

$$\Pi(\theta) = Ae^{-\frac{U}{k_B T}} = Ae^{\frac{\mu_J B \cos \theta}{k_B T}}, \quad (9.30)$$

where k_B is the Boltzmann constant. In this case as well, as for the polar dielectrics, the exponent is generally quite small, and we can expand the exponential in series stopping at the first term. The argument continues exactly as that in Sect. 4.7 for the polar dielectrics, and we will not repeat it. The conclusion is that

$$\mathbf{M} = n_p \langle \mu \rangle = \frac{n_p \mu_J^2}{3k_B T} \mathbf{B}. \quad (9.31)$$

where μ_J is the intrinsic magnetic moment and n_p is the number of molecules per unit volume.

We have found that the induced magnetization is in the direction and sense of the inducing field and is proportional to it. The simple proportionality is a consequence of the smallness of the exponent on Eq. (9.30), a feature that has allowed us to replace the exponential with the first term of its development in series. Equation (9.31) is called Curie's law, from Pierre Curie (France, 1859–1906), who established it in 1895. The law also tells us that \mathbf{M} is inversely proportional to the absolute temperature. Contrastingly, diamagnetism is independent of temperature.

In our model, the magnetic susceptibility is

$$\chi' \approx \chi = \frac{n_p \mu^2 \mu_0}{3k_B T}, \quad (9.32)$$

which is a positive quantity. Susceptibilities of paramagnetics are usually small, but somewhat larger than for diamagnetics. Some examples are in Table 9.2. The elements with one unpaired electron, namely the alkalines (Na, K, Rb, Cs), and Mg,

Al and Mn, are paramagnetic. Oxygen is a particularly strong paramagnet. Note that the magnetic moment induced by deformation of the atoms is present in the paramagnetic substances too, but is usually subdominant compared to the orientation effect.

QUESTION Q 9.1. Using the information found in this chapter, what is the magnetic susceptibility of aluminum at 200 K? \square

9.7 Ferromagnetism

As we have already stated, in a ferromagnet, the relations between the \mathbf{B} , \mathbf{H} and \mathbf{M} fields are not linear. We shall see now that they are not even single-valued. Let us consider the following experiment. Begin with the setup shown in Fig. 9.17, consisting of an iron torus around which we have tightly wound N turns of a conductive wire, a battery, a variable resistor and an amperometer. The battery will produce a current of constant intensity I that we can regulate acting on the variable resistor and measure with the amperometer.

The \mathbf{H} field can be simply calculated using the Ampère law. Symmetry tells us that the lines of \mathbf{H} are circles concentric with the torus. We then choose, as a closed curve of the Ampère law, one of these circles, say, Γ inside the solenoid, and we write

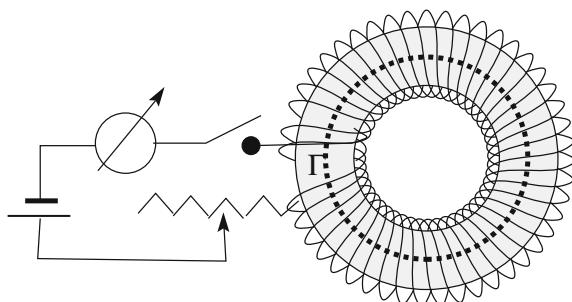
$$\oint_{\Gamma} \mathbf{H} \cdot d\mathbf{s} = I_{\text{linked}}. \quad (9.33)$$

Equation (9.33) immediately gives us $Hh = NI$, where h is the length of Γ and

$$H = NI/h. \quad (9.34)$$

We see that H is inversely proportional to the total length of the *magnetic circuit*. In addition, H is directly proportional to the product of the current intensity times the number of turns. This quantity is called the *magnetomotive force* and is

Fig. 9.17 The set-up for measuring the \mathbf{B} field as a function of the \mathbf{H} field



measured in *ampere-turns*. This conclusion is quite general, as we shall discuss in Sect. 9.8, and consequently, the measurement unit for H is often called *ampere-turn per meter*. Obviously, this is nothing more than an ampere per meter (A/m.), given that the number of loops is dimensionless. In general, we can directly control the field \mathbf{H} by controlling the currents.

If we now want to know \mathbf{B} , we would need a relation between \mathbf{B} and \mathbf{H} , or a relation between \mathbf{B} and \mathbf{M} . As a matter of fact, such single-valued relations do not exist, because the magnetization \mathbf{M} does not simply depend on the actual value of \mathbf{B} , but also on its precedent values, or, as it is said, on the magnetic history of the material.

Let us use our setup and start measuring. We have taken care to demagnetize the iron by heating it to a very high temperature, where the ferromagnetic properties are lost (as we shall soon see). This process cancels the magnetic history. We now want to measure H and B at different values of the current. The former is immediately given by Eq. (9.34) measuring I . As for \mathbf{B} , symmetry tells us that its lines are circles concentric to the torus as well. We also know that it is zero outside the solenoid. We can measure \mathbf{B} in the iron by opening a small slot on the torus perpendicular to its axis. \mathbf{B} being continuous across a surface to which it is normal, \mathbf{B} in the gap is equal to \mathbf{B} in the iron.

In Sect. 7.5, we saw how to measure \mathbf{B} with the flip-coil method, taking advantage of Felici's law. We can use a similar method introducing a small coil into the gap, made of winding n loops (to enhance any effect) of area S . We cannot flip it by 180° but we can extract it quickly and measure the total charge Q . We lose a factor of 2 compared to Eq. (7.21). The flux variation is equal to the flux inside, which we call Φ_{in} , because the flux outside is zero. Equation (7.21) without the factor 2, if R is the resistance of the measuring circuit, gives us

$$Q = \Phi_{\text{in}}/R = nSB/R, \quad (9.35)$$

from which we know B , the quantities n , S and R being known by construction (or, even better, by calibration).

Finally, by measuring the fields B and H and varying the current intensity, we obtain the diagram in Fig. 9.18, which is called the *magnetization curve*. Let us discuss it.

The diagram in Fig. 9.18 has the magnitude B on the ordinate axis and $\mu_0 H$ rather than H on the abscissa, in order to have homogeneous quantities on the two axes. Notice that the $\mu_0 H$ scale is about four orders of magnitude wider than the B scale.

Curve 1 shows the performed measurements starting, as we have said, after having canceled the magnetic history. One sees that when H increases, B increases much more, initially almost linearly, but very soon non-linearly. Already, at quite small values of $\mu_0 H$, the values of B are very large. Notice that $\mu_0 H$ is what the value of B would have been if the solenoid had contained air instead of iron. Hence,

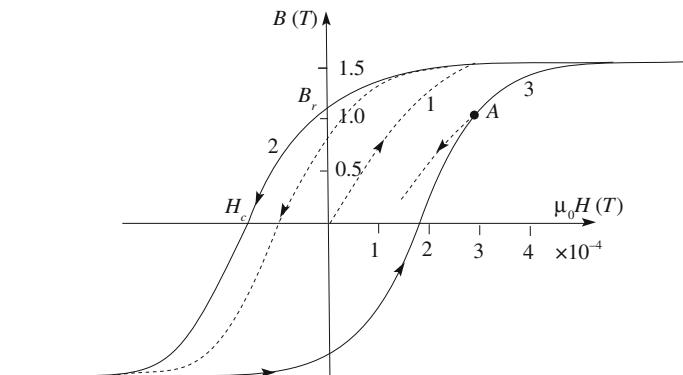


Fig. 9.18 Magnetization curves

it is the iron that produces such a high field. The fact is that even modest values of H induce a very large magnetization M . The latter is equivalent to a huge current density on the surface of the torus. As we know, B is due to both conduction and magnetization currents.

If we further increase H , the curve tends to level, or, as it is said, to saturate, as shown in Fig. 9.18. In the scale of the diagram, it looks like B would become a constant. This is not really true. Rather, B continues to rise for increasing $\mu_0 H$, even if slowly. The slope of this part of the magnetization curve is equal to one. The saturation value for iron (depending on the alloy) is around 2 T. To be precise, the saturation is not of B , but of M , which cannot be larger in practice than a certain value.

Let us now suppose that, once we have reached saturation, we begin gradually decreasing the current, and hence H , always measuring B . We find out that our measurements do not follow, in the opposite direction, the initial curve, but rather we are now on branch 2 in the figure. As anticipated, B is not a single-valued function of H .

Notice, in particular, that at zero current, namely when $H = 0$, B is not zero but may have a sizeable value. This is due to a residual magnetization, which we call M_r . The torus is now permanently magnetized. Let us, however, continue, sending a current of the opposite sign. We are now increasing H along the negative axis of the diagram. We are still on branch 2. When H reaches the value marked as H_c , called *coercivity* and also the *coercive field*, the B field is zero. Notice that the magnetization is not zero at this point, but rather, according to Eq. (9.8), is equal to $-H_c$. If we continue to increase the negative value of H along prong 2 of the curve, we reach saturation again, with inverted signs.

If we now invert the process once more, namely we gradually take the current to zero, invert it and increase over positive values, we find branch 3 in Fig. 9.18. The set of curves 2 and 3 together is called a hysteresis cycle.

We can still learn more. Suppose, for example, we move along branch 3 increasing the current, but we do not go as far as saturation. Rather, we stop at point A in the figure, and then decrease the current, and H . We find that we now move along a new branch, the initial part of which is shown dotted in the figure. With similar actions, we can reach every point inside the two extreme curves 2 and 3.

In conclusion, for every value of H , the magnetic field B , and the magnetization as well, can take an entire interval of different values. The actual value of that interval depends, as we anticipated, on the magnetic history, namely what has happened before.

We have not yet seen how to demagnetize the metal, namely to reach the point $H = B = 0$ in the diagram. There are two ways. We can, as already mentioned, heat it above the transition to non-ferromagnetic temperature, or we can go through a series of hysteresis cycles of decreasing amplitude.

Notice that the shape of the magnetization curves, of which Fig. 9.18 is an example, strongly depends on the material, in particular, on the alloy.

Hysteresis is a dissipative phenomenon, corresponding to a loss of macroscopic energy (which transforms into thermal energy). Indeed, let us suppose gradually changing the current, similarly to what we did above, going through a cycle in the (H, B) plane. As B varies with time, so does its flux linked to the solenoid. Consequently, the emf $\mathcal{E} = -ANdB/dt$ appears at the ends of the solenoid (A is the area of a loop and N is the number of loops). Neglecting the passive resistances, in order to have the current intensity I , the battery must spend a power, namely energy per unit time, equal to $dU/dt = -\mathcal{E}I = HhAdB/dt$, where we have used Eq. (9.34) to express I as a function of H .

The total energy spent is obtained by integration over the cycle, namely

$$U = hA \oint HdB, \quad (9.36)$$

which is the area of the cycle in the (H, B) plane times the volume of the torus.

From what we have discussed, it is clear that alloys with the narrowest possible cycles (namely with the smallest area) must be used in the transformers nucleuses, while alloys with the widest cycles (namely with the largest residual magnetization) must be used for permanent magnets. \square

QUESTION Q 9.2. You use the method mentioned above to measure a field intensity $B = 0.2$ T moving a coil of 1 cm^2 area made of 50 turns. The internal resistance of your ballistic galvanometer is 100Ω . What is the charge you measure? \square

QUESTION Q 9.3. Consider a magnetic circuit 40 cm long, with a section of 10 cm^2 , having the hysteresis loop in Fig. 9.18. Approximate the loop with a parallelogram and evaluate the energy dissipated in 100 loops. \square

9.8 Uses of Ferromagnetism

As we have seen, in ferromagnetic materials, relative modest conduction currents are sufficient to produce high values of B . This property makes ferromagnetics very useful in the construction of transformers, electrical motors, electromagnets, etc. Note that periodic variations with time of the conduction currents in the coils drive the material through cycles of hysteresis. This implies losses of macroscopic energy, proportional to the area of the cycle.

As a first example, we consider the use of ferromagnetism to build inductors of high inductance. Inductors are made by winding a wire in a (large) number of loops. To increase its inductance, we can do that around a small ferromagnetic ring. The geometry is like that in Fig. 9.17, but on a smaller scale. If we remember that the inductance is the ratio between the flux of B and the current intensity that is its cause, we see that an iron core, as it is called, may increase the inductance by large factors. Let us look more closely at the issue, considering that the inductors are used in electronic circuits in which the current intensities are generally quite small. Take, for example, the first segment of curve 1 in Fig. 9.18, where we can consider the dependence of B on H to be linear, at least in a first approximation. We can then write

$$B = \mu H = \mu_0 \kappa H = \mu_0 \kappa N I / h. \quad (9.37)$$

The values of the magnetic permeability relative to vacuum κ , in the limits in which it can be defined, are very high for ferromagnetics, ranging from several thousands to millions. κ tells us how many times B is larger than $\mu_0 H$, which is the value it would have in a vacuum.

We calculate the inductance L of our torus starting from the flux of \mathbf{B} linked to the current, which is $\Phi_B = NBA = \mu_0 \kappa I A N^2 / h$, where A is the area of the loops (namely the cross section of the torus) and h is the length of the torus. The inductance is then

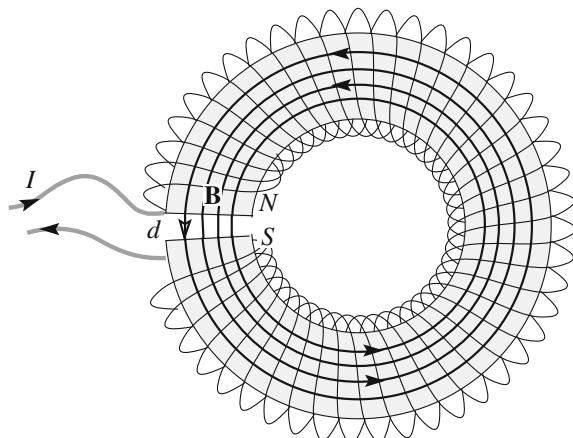
$$L = \kappa \mu_0 N^2 A / h. \quad (9.38)$$

As expected, the inductance is proportional to κ , which, as mentioned, can easily be 10^3 – 10^4 . By the same token, the usefulness of iron nuclei in transformers and electrical motors should be clear.

QUESTION Q 9.4. The alloys commonly used to manufacture inductors are the ferrites. You have a small torus of a ferrite with $\kappa = 640$, which is 1 cm long and has a 4 mm^2 section and a sufficiently long copper wire. If you want an inductor of $L = 1 \text{ mH}$, how many turns do you have to wrap? \square

As a second example, let us consider electromagnets. Several cases exist in which an object must be introduced into a strong magnetic field. Think, for example, of NMR apparatuses, into the field of which parts (or all) of the human body must be introduced for medical analysis, to the magnets used to deflect or guide particles or nuclear beams, to a cyclotron, or to a mass spectrometer.

Fig. 9.19 Schematic view of an electromagnet



Electromagnets are used to produce the magnetic field of requisite intensity in the useful space region, as we already did on several examples. We can start again from Fig. 9.17 (the torus must now be thought of as being much larger and heavier than for an inductor). The core must be interrupted to make the useful space available. Let us cut a short gap, of thickness d , as in Fig. 9.19, creating two magnetic poles, a North one and a South one. Such systems are called dipole magnets.

If d is sufficiently small and in a first approximation, the \mathbf{B} and \mathbf{H} fields in the torus have not changed. As for the fields in the gap, consider that the surfaces of the cut are normal to the lines of \mathbf{B} . Being that the normal component of \mathbf{B} continuous, \mathbf{B} in the gap is not very different from \mathbf{B} in the iron. To be precise, its lines open a bit at the borders, similarly to the lines of \mathbf{E} at the borders of a parallel plate capacitor, but the effect is small. As matter of fact, the magnetic flux Φ_B is equal through all the sections, both in the iron and in the gap. For this reason, we talk of a *magnetic circuit*.

The behavior of \mathbf{H} is different. Indeed, when we cut the torus, we produced two surfaces, namely the poles, on which $-\nabla \cdot \mathbf{M}$, i.e., the magnetic charge density, which we call ρ_m , has opposite signs and is large. We know that the circulation of \mathbf{H} about a closed line is equal to the linked conduction current. Let us choose a circle, Γ , inside the ring. Symmetry tells us that the field of \mathbf{H} is always tangent to Γ and has two different values, H_F in iron and H_G in the gap. The Ampère law gives us

$$H_F h + H_G d = NI,$$

where we have considered the length of the curve inside the iron to be h , rather than $h-d$, given that $d \ll h$. Now, in the gap $H_G = B/\mu_0$, where B is the magnetic field, both in the iron and in the gap. We can then write the last equation as

$$(\mu_0 H_F)h + Bd = \mu_0 NI, \quad (9.39)$$

The equation contains two unknowns, H_F and B . To solve it, we need a relation between H_F and B . However, as we know, there is no single-valued relation between these two quantities. In other words, for a given current intensity, B , and hence H in the gap, are not uniquely defined, but rather depend on the magnetic history. We discussed this issue in Sect. 9.7, where we also saw how experimentally to determine the relation between H_F and B .

Suppose that we have obtained the hysteresis cycle in Fig. 9.20 with that procedure. Once we have it, we can solve the problem graphically. We notice that in the plane $(\mu_0 H_F, B)$, Eq. (9.40) represents a straight line, having slope $-h/d$, and intercepts $\mu_0 NI/h$ and $\mu_0 NI/d$ on the horizontal and vertical axes, respectively. Note that the slope is independent of the current intensity. When I varies, the line moves parallel to itself.

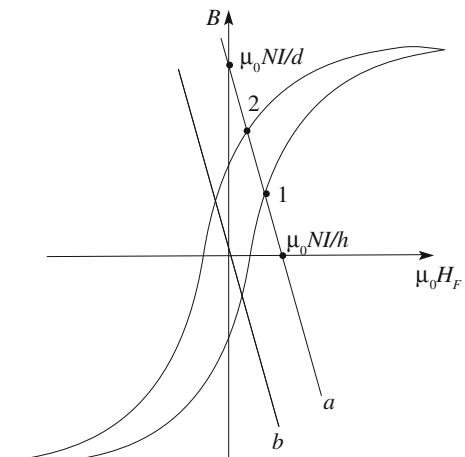
Consider the line (a) in Fig. 9.20 relative to a generic value I of the current intensity. The solution is given by the point at which the line cuts the curve inside the hysteresis cycle we are moving on. It can be any between points 1 and 2.

Particularly interesting is line b , which is for $I = 0$. The points where it cuts the magnetization curve give the values of H_F (inside the iron) and B (both inside the iron and in the gap). As we see, the field H_F is not necessarily zero.

What is zero is the circulation of H , and now, in contrast to when the iron torus is complete, the values of H are different in the iron and in the gap. Indeed, they have opposite signs.

If the hysteresis cycle is sufficiently narrow and we are looking for an approximate solution for small values of B , we can solve the problem analytically assuming that $B \simeq \mu H_F$. Graphically, this means approximating the cycle with a straight line. In this case, we can substitute $H_F = B/\mu$ in Eq. (9.39), obtaining

Fig. 9.20 The hysteresis cycle and Eq. (9.40) for a generic current intensity (a) and null current (b)



$$B = \frac{NI}{h/\mu + d/\mu_0}. \quad (9.40)$$

As we have noted above, the flux of B , Φ_B , is constant along the entire magnetic circuit. To obtain its expression, we multiply Eq. (9.40) by the section area A (which is also constant in this case), obtaining

$$\Phi_B = \frac{NI}{\mathcal{R}}. \quad (9.41)$$

with

$$\mathcal{R} = \frac{h}{A\mu} + \frac{d}{A\mu_0}. \quad (9.42)$$

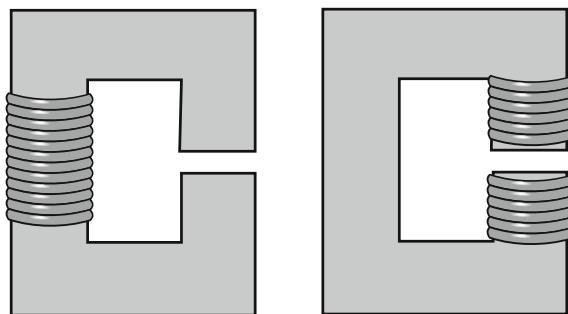
Hence, the magnetic flux is equal to the ratio between the *magnetomotive force*, namely the number of *ampere-turns* NI , and the *magnetic reluctance* \mathcal{R} . The latter is the sum of the reluctances of the parts in series of the magnetic circuit, one in iron and one in air. The reluctance of each segment is directly proportional to its length inversely to its section and to the magnetic permeability of the medium.

Equations (9.41) and (9.42) are very similar to Ohm's laws. They are called Hopkinson's laws, after John Hopkinson (UK, 1849–1998), but also Ohm's laws for magnetic circuits. Hopkinson's laws hold only in a first approximation, but are often very useful in practice in the design of magnetic circuits for a preliminary evaluation of the orders of magnitude of the problem.

QUESTION Q 9.5. Consider a ferromagnetic torus with a gap, as in Fig. 9.19. It is 2 m long, has a section that is $A = 800 \text{ cm}^2$ and the gap is 2 cm wide. The iron has the initial magnetization curve shown in Fig. 9.18. Find the reluctances of the core and of the gap. Repeat with a 5 cm wide gap. \square

The configuration of Fig. 9.19 has been useful in its simple geometry for our discussion of the working principles of electromagnets. The design of electromagnets for practical uses is a full branch of applied physics. Electromagnets come in very different shapes. However, in any case, the major part of the magnetic circuit is in iron, called the iron core. The useful part is the gap in air, whose reluctance is $\mu \gg \mu_0$. Consequently, the reluctance of the gap is much larger than that of the part in iron. The length of the non-iron parts should be reduced to a minimum, but the gap should still be wide enough to serve its purpose. Once the geometry has been defined, the magnetic flux is proportional to the ampere-turns. Given this, the coils can be located where it is most practical, not necessarily distributed all around the torus, as in Fig. 9.19. In addition, the iron core does not need to be a torus, but can have any practical shape, provided its cross-section is always large enough so as not to increase the reluctance [see Eq. (9.42)]. Figure 9.21 shows two examples of common geometries.

Fig. 9.21 Two common geometries of an electromagnet



9.9 Microscopic Interpretation of Ferromagnetism

As we have already mentioned, atomic electrons are usually paired into couples in which the two spins are in opposite directions. The total spin of the pair is zero. Ferromagnetic materials, like paramagnetic ones, have atoms with one or two unpaired electrons. Consequently, their magnetic spin moments may contribute to the macroscopic magnetization \mathbf{M} . The characteristic property of ferromagnetics is that the unpaired electrons of different atoms (and of the same atom if they are more than one) have a strong tendency to align parallel with one another. This tendency is not limited to distances of the atomic scale (which, remember, are tenths of a nanometer), but extends over many micrometers or more, which are enormous distances on the atomic scale. The interaction responsible for the alignment cannot be due to the magnetic interaction between magnetic moments, because its intensity is by far too large. Once more, the nature of this interaction is purely quantistic. Its consequence is that, in ferromagnetic materials, the magnetic moments of the unpaired electrons spontaneously align in the same direction in the absence of an external field \mathbf{B} .

Many ferromagnetic materials exist. Some of them are pure elements (Ni, Fe, Co, Gd), many more are alloys. The ferromagnetic substances are in the solid phase, namely they are crystals. Their ions have equilibrium positions at the vertices of a regular lattice. For example, the unit cell is a cube for iron, a face-centered cube for nickel, and a hexagonal prism for cobalt. Usually, the regular structure is not seen by the naked eye, because the material is polycrystalline, namely made by micro-crystals, whose dimensions are typically between 1 and 100 μm .

The anisotropic structure tends to align the magnetization in a small number of preferred directions (along the sides and/or the diagonals of the unit cell). These directions are called *easy magnetization axes*. For example, iron has six easy directions, nickel has eight of them, and uniaxial crystals have only one easy direction. The spontaneous magnetization process leads, in any case, to the configuration of minimum energy. Consider, for simplicity, a uniaxial crystal. Even in this simplest case, energy can be minimized with a partition of the ferromagnet into a number of sub-divisions. Indeed, if the whole crystal is magnetized in the easy

direction in the same sense, as in Fig. 9.22a, the magnetic field outside the crystal extends over a considerably large volume, and correspondingly, the magnetic energy stored in its volume is large. If the magnetization in the two halves of the crystal has the same direction but opposite senses, as in Fig. 9.22b, the energy of the external magnetic field is smaller. If the regions of magnetization opposite to one another are four, as in Fig. 9.22c, the field energy is even smaller.

However, the field energy does not represent all the energy in the game. Consider the separation surface, called the magnetic walls, between two regions, or domains, of opposite magnetization. The spins on the two sides of the wall have opposite senses. The interaction responsible for the alignment would tend to align them. This means that the energy of parallel spins is smaller than the energy of anti-parallel spins. Hence, introducing a wall increases the energy of the system by a *boundary energy*, which is proportional to the surface of the wall. When the number of domains in the crystal increases, the boundary energy, which is a surface energy, increases, while the field energy, which is a volume energy, decreases. Consequently, a configuration exists in which the sum of the two is a minimum. This is the equilibrium configuration in the absence of external agents. Each crystal contains a number of regions of uniform magnetization, called *Weiss domains*, after Pierre-Ernest Weiss (France, 1865–1940), who developed the domain theory in 1907.

Consider now the polycrystalline structure. The easy orientations of the microcrystals are distributed at random, as shown schematically in Fig. 9.23. This explains why the magnetization does not usually appear in a macroscopically-sized ferromagnet, such as a common piece of iron, even if it is present at the microscopic level. Notice, however, for the sake of precision, that the cold rolling process used to produce iron strips, of wide commercial application, tend to align the micro-crystals, which consequently have an easy direction parallel or near to the rolling direction.

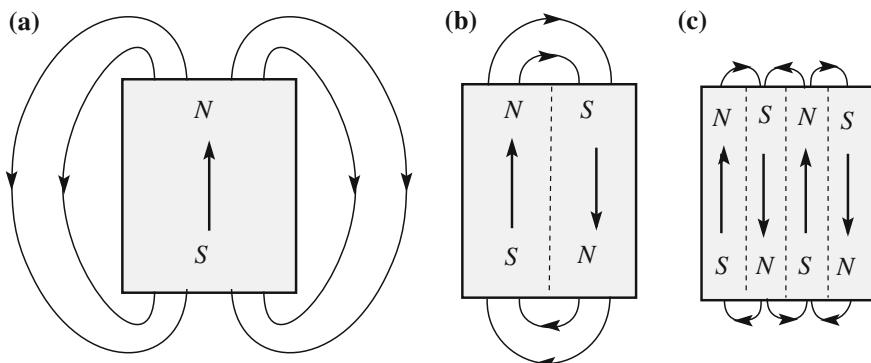
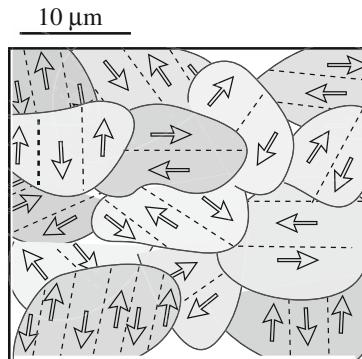


Fig. 9.22 Schematic representation of a ferromagnetic microcrystal with the magnetic field produced by magnetization in the surrounding space. The longer sides are parallel to the easy axis. **a** A single domain, **b** two domains, **c** four domains

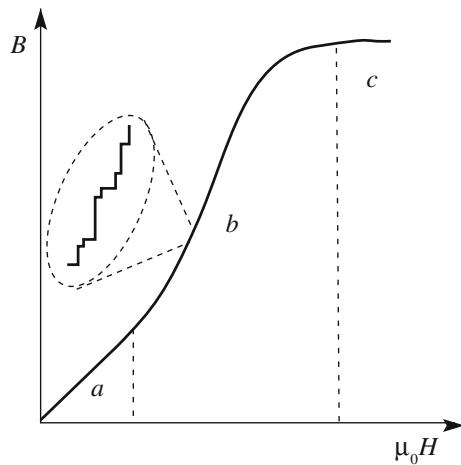
Fig. 9.23 Sketch of the crystallites and Weiss domains in a ferromagnet. Note that the easy magnetization direction does not always belong to the plane of the figure



Let us now take a demagnetized piece of iron through the magnetization curve, introducing it into a current-carrying solenoid. Let us first think of the microcrystals with an easy axis parallel (or almost so) to field \mathbf{B} in the solenoid. The energy of one of these crystals will diminish if the domains with magnetization \mathbf{M} in the sense of \mathbf{B} grow at the expense of those with opposite \mathbf{M} . In fact, a domain grows with a movement of its walls toward the outside. For small values of H , the displacement is small and the process is reversible. Namely, if we diminish the applied field rather than increase it, each wall inverts its motion. This corresponds to part (a) of the curve in Fig. 9.24.

For higher values (part b of the curve), the motion of the walls becomes irreversible. When a wall reaches a crystal defect, it sticks to it, so to speak, and stops up to when the field has grown enough to tear it off. Once free of the obstacle, the wall suddenly jumps to the new equilibrium position, corresponding to that value of the field. These are called Barkhausen jumps, after Heinrich Barkhausen (Germany, 1881–1956), who discovered the process in 1919. Energy is lost in a

Fig. 9.24 An initial magnetization curve



jump, because the rapid local variations of the magnetic field produce eddy currents, whose energy is lost through the Joule effect. The insert in Fig. 9.24 shows an enlarged segment of the resulting magnetization curve. If we proceed with a decreasing applied field, the “jumping motion” does not follow the exact same path in the inverted direction. The process is irreversible.

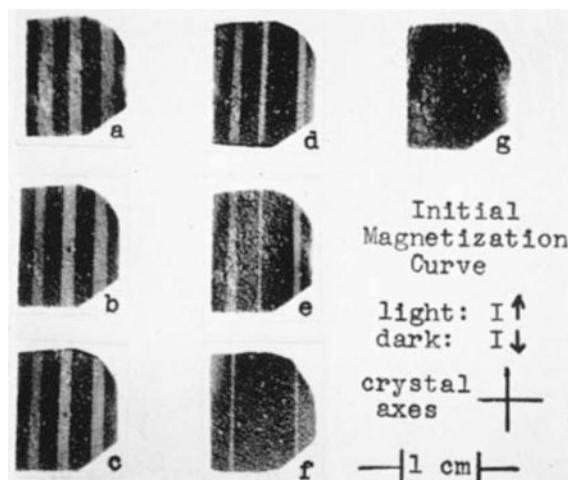
Several techniques exist to visualize the magnetic domains. One of them consists of illuminating the sample with polarized light and observing it in reflection through a polarization filter (see Volume 4). The initial state of light polarization is altered by the magnetic field on the surface in opposite ways for opposite senses of the magnetization. Consequently, the images of adjacent domains of opposite \mathbf{M} appear to be of contrasting intensity. Without entering into any detail, we simply show, in Fig. 9.25, a sequence of pictures of the face of a single crystal of a Si–Fe alloy. The crystal was quite large, namely a slab of $17 \times 14 \times 3 \text{ mm}^3$, cut to have an easy axis on the face. Picture (a) was taken after demagnetization of the sample, and the subsequent ones, from (b) to (g), for increasing H -field applied parallel to the easy axis, up to complete saturation (in g). One sees how the “black” domains become larger and larger, at the expense of the “white” ones, until, when saturation is reached, the whole crystal has become a unique Weiss domain.

Let us now go back to our polycrystalline structure. What about the micro-crystals whose easy axes are not parallel to the applied field? Well, they do the best they can, namely as the domain at the smaller angle to the field grows, the other becomes smaller.

For very high fields, we reach the segment c of the curve. Here, the whole crystallites are single Weiss domains. Increasing the applied field, we do, in fact, still increase somewhat the magnetization \mathbf{M} . This is done by turning in the crystals whose easy axes are at angles with the applied field. This is a very hard process, however, and consequently, the slope of the curve c is very small.

Fig. 9.25 The opposite sign domains on the face of a SiFe crystal (see text) for growing applied field. Reprinted with permission from C.H. Fowler and E.M. Fryer; Physical Review 94 (1953) 52.

Copyright (1953) by the American Physical Society (<http://journals.aps.org/pr/abstract/10.1103/PhysRev.94.52>)



A simple and spectacular demonstration of the domain orientation process in the central (*b*) part of the magnetization curve was made by Heinrich Barkhausen in 1919. It is called the Barkhausen effect and is shown in Fig. 9.26.

You should prepare several iron sticks (you can use the iron used for transformer cores, for example) and introduce them into a solenoid of a few thousand loops. Any electromotive force induced in the solenoid is amplified by an electronic amplifier, whose output is sent to a loudspeaker. We now move a permanent magnet near to the sticks. The field of the magnet causes the orientation of a certain number of domains, namely the movement of a number of walls. The corresponding flip of atomic magnetic moments produces a small variation of the flux **B** linked to the circuit, which generates an emf. After amplification, we hear the emf as a “crick”. When we move the magnet along the direction of the sticks, we hear a sequence of such cracks. We are listening to the moving atomic magnetic moments!

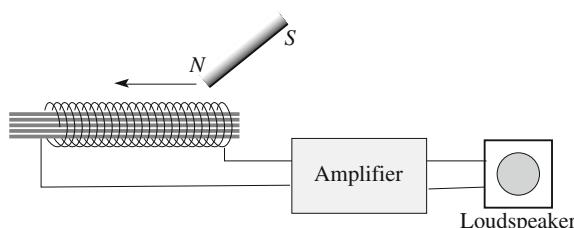
When we have reached the end of the sticks, we move the magnet a short distance away, bring it back to the end from which we started, and repeat the previous movement. The noise is now much fainter. Repeating the process again, the noise soon disappears completely. All the domains have been oriented. If we now move the magnet back in the opposite direction (or in the same direction with the other pole of the magnet near to the sticks), the noise is strong again, because we are orienting the domains in the direction opposite to that which they have.

As a final observation, we can check the reversibility of the process for small disturbances. We put the pole of our magnet near a point of the sticks and then move it back and forth by a small distance. The noise we hear is very faint; the process is locally reversible.

We conclude with an important observation. Ferromagnetism is a particular phase of the ferromagnetic materials, which exists only below a well-defined critical temperature, a characteristic of the substance, or alloy. It is called the Curie temperature, T_c , after Pierre Curie. Above its Curie temperature, the material is paramagnetic.

Curie temperatures for the most common, ferromagnetic pure elements are 770 °C for Fe, 1131 °C for Co, and 358 °C for Ni. The ferromagnetic-to-paramagnetic transition can be easily verified. For example, a wire of nickel (whose Curie temperature is relatively low) is attracted by a magnet at room temperature. If we heat it on a flame above T_c , the magnet does not attract it anymore. If we wait a bit, we see the wire jumping to the magnet pole when its temperature has sufficiently decreased.

Fig. 9.26 Demonstration of the Barkhausen effect



Gadolinium has an even lower Curie temperature, namely $T_c = 16$ °C. It is usually paramagnetic at room temperature, but you can induce the transition to ferromagnetic by simply dropping it into a glass of water with some ice in it. This allows for performing amusing experiments and tricks (think of them yourself and have a look on the web).

9.10 Energy of Steady Currents in the Presence of Magnetic Materials

In Sect. 8.4, we discussed the energy of a system of steady (macroscopic) conduction currents in a vacuum. The system is defined when the current density $\mathbf{j}_c(\mathbf{r})$ is known for every \mathbf{r} in space. We have also seen that the energy is stored in the magnetic field \mathbf{B} generated by $\mathbf{j}_c(\mathbf{r})$ as

$$U_m^0 = \frac{1}{2\mu_0} \int_{\text{all space}} B^2 dV. \quad (9.43)$$

When materials are present, the current density is the sum of two terms: the conduction current \mathbf{j}_c and the microscopic magnetization currents. We want here to express energy in terms of the macroscopic currents only. Under certain aspects, the situation is similar to that considered in Sect. 4.9, where we discussed the analogous problem in electrostatics, considering the free charges energy. However, the two cases also have important differences. We recall that we are considering the macroscopic magnetic field, namely the field mediated on distances and times substantially larger than the atomic ones. We also recall that, being that the energy is proportional to the square of the field, the macroscopic field energy is different from that of the microscopic field.

Let us start from a geometrically simple case. We consider a straight solenoid, whose length l is much larger than the diameter of its section S . Let I be the current intensity and n the number of turns per unit length. The solenoid is completely full of the material under analysis. Suppose that the \mathbf{H} field is proportional to \mathbf{B} , namely that $\mathbf{B} = \kappa\mu_0\mathbf{H}$. Here, κ is the magnetic permittivity, which we shall assume to be constant. The material may be paramagnetic (or ferromagnetic at very small magnetization), and then $\kappa > 1$, or diamagnetic, and then $\kappa < 1$. Let L be the inductance of the solenoid and U_m^0 its magnetic energy in absence of the core (namely in a vacuum). The energy of the system with the core is

$$U_m = \frac{1}{2}LI^2 = \frac{1}{2}\kappa n^2 I^2 \mu_0 l S = \kappa U_m^0. \quad (9.44)$$

Notice that $U_m > U_m^0$ if the material is paramagnetic or ferromagnetic, $U_m < U_m^0$ if it is diamagnetic.

The real solenoids always have a finite length, and consequently, their magnetic field is different from zero outside, especially near the faces. This fringe effect is important in the subsequent arguments. We made the hypothesis of a very long solenoid just to find, in a easy way, that $U_m = \kappa U_m^0$, an expression that we shall see will hold, in general.

As always, the energy of a state is the work to be done to bring the system into that state from the state of zero energy. We take as a zero energy state the state in which the solenoid is empty and the current is zero. To reach the desired state, we can proceed as follows. We switch on the current up to the desired value, reaching the state shown in Fig. 9.27a. The (electric) work done on the system, by the battery, is U_m^0 . We now introduce the core. When the core is in the non-homogeneous field region, it is under the action of a force, which is attractive toward the inside in the paramagnetic case, ($\kappa > 1$), repulsive toward the outside in the diamagnetic case ($\kappa < 1$). Let us discuss the paramagnetic case (the diamagnetic case is completely analogous), shown in Fig. 9.27b. The (mechanical) work to be done on the system is the work toward the magnetic force \mathbf{F}_m shown in the figure. This work is negative. One might think, at this point, that the final energy would be smaller than U_m^0 . However, we have just seen that $U_m > U_m^0$. What did we do wrong?

We did nothing wrong; rather, we have not finished yet. Indeed, in the process of introducing the core, the current intensity must remain constant, and this does not happen without spending work. Indeed, during the movement of the core toward the inside, the field \mathbf{B} increases and its flux linked to the solenoid increases as well. This generates an emf that, for the Lenz law, opposes the cause that has generated it, namely it tends to diminish the existing current. Consequently, we must keep the battery connected while moving the core in. The battery will spend a positive electric work on the system. We shall not develop the calculation here, but only state that the result is that the sum of the mentioned mechanical work (negative) and electric work of the battery (positive) is indeed positive and equal exactly to $U_m - U_m^0$.

The case of the diamagnetic core is completely analogous, once the signs are properly changed.

Let us now go back to the magnetic energy given by Eq. (9.44). We see that we can write it as

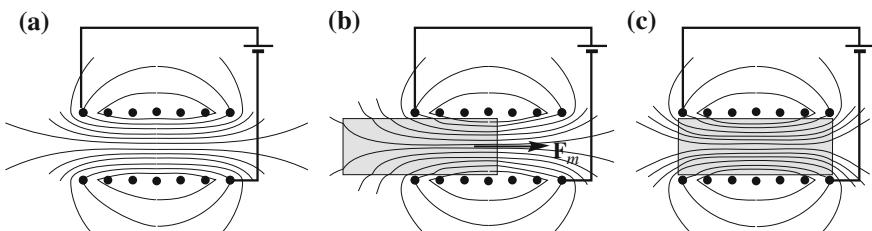


Fig. 9.27 **a** A solenoid in a vacuum, **b** introducing the magnetic material core, **c** the core is completely inside

$$U_m = \frac{1}{2}(nI)(\kappa\mu_0 nI)(lS) = \frac{1}{2}\mathbf{H} \cdot \mathbf{B}V. \quad (9.45)$$

In this particular case, at least, energy appears to be distributed in space with density

$$w_m = \frac{1}{2}\mathbf{H} \cdot \mathbf{B} \quad (9.46)$$

and total value

$$U_m = \frac{1}{2} \int_{\text{all space}} \mathbf{H} \cdot \mathbf{B} dV. \quad (9.47)$$

We shall now show that these equations hold, in general, for *linear* magnetic materials, namely when $\mathbf{B} = \kappa\mu_0\mathbf{H}$, under time-independent conditions.

Our arguments will be quite similar to those of Sect. 8.4. As in that discussion, we consider a system of steady conduction currents $\mathbf{j}_C(\mathbf{r})$ in the volume V , but now in the presence of magnetic materials. We start from the state in which all the conduction currents are zero. The magnetization currents are zero as well (as we are not considering permanent magnets). We now gradually increase the conduction currents. The materials magnetize, namely they acquire magnetic dipole moments or, in other words, magnetization currents are produced, which grow with time. The time variation of the magnetic field \mathbf{B} of both currents generates an electric field \mathbf{E} according to

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}. \quad (9.48)$$

The (electrical) work to be done on the charges of the conduction currents against the induced field in the infinitesimal time interval δt is $-\mathbf{E} \cdot \mathbf{j}_C \delta t$ per unit volume. Integrating on the total volume of the system of currents V , we have

$$\delta W = -\delta t \int_V \mathbf{E} \cdot \mathbf{j}_C dV = -\delta t \int_{\text{all space}} \mathbf{E} \cdot \mathbf{j}_C dV,$$

where, on the right-hand side, we extended the integral to the entire space. This is allowed, the integrand being zero outside V , and will turn out to be useful, as we shall soon see.

We now eliminate \mathbf{j}_C using the equation $\nabla \times \mathbf{H} = \mathbf{j}_C$, obtaining

$$\delta W = -\delta t \int_{\text{all space}} (\nabla \times \mathbf{H}) \cdot \mathbf{E} dV.$$

We proceed with the same arguments as in Sect. 8.4, obtaining in place of Eq. (8.16)

$$\delta W = \delta U_m = \int_{\text{allspace}} \mathbf{B} \cdot \delta \mathbf{H} dV. \quad (9.49)$$

To go further, we need a relation between \mathbf{H} and \mathbf{B} . We shall restrict ourselves to the linear materials, in which the two field magnitudes are proportional to one another. Then, and only then, we have $\mathbf{B} \cdot \delta \mathbf{H} = \delta(\mathbf{B} \cdot \mathbf{H})/2$, and we can write

$$\delta U_m = \frac{1}{2} \int_{\text{allspace}} \delta(\mathbf{B} \cdot \mathbf{H}) dV.$$

The total work to be done on the system is the integral of δU_m from the state in which \mathbf{B} and \mathbf{H} are zero to the state in which they have their final values. The result is Eq. (9.47), as we wanted to prove.

Summarizing, Eqs. (9.46) and (9.47) express the magnetic energy of a system of steady currents in the presence of linear magnetic materials. In other words, these equations hold for para- and dia-magnetic materials. They also approximately hold for ferromagnetic materials at very low values of the fields starting from demagnetization.

We note that, in any case, the magnetization process implies rearrangement and reorientation of the molecules. We are dealing with a very complex process. The linearity hypothesis leads to simple expressions, as in the electrostatic case. When the hypothesis does not hold, the energy of the system must be calculated by integrating Eq. (9.49). This may be very difficult, as one might understand, remembering that the result may depend on the history of the material, namely in the presence of hysteresis.

Summary

In this chapter, we learned the following principal concepts:

1. Magnetic materials: dia-, para, and ferro-magnetic
2. The magnetization vector field
3. The relation between magnetization and the magnetization (microscopic) current.
4. The vector field \mathbf{H} and its causes
5. The equations ruling the time-independent magnetic field in the presence of materials.
6. Magnetic permeability and susceptibility
7. The properties of the \mathbf{B} and \mathbf{H} fields

8. The field \mathbf{H} in absence of macroscopic currents
9. The non-existence of a magnetic charge
10. Microscopic interpretation of dia- and para-magnetism
11. The magnetic hysteresis
12. Microscopic interpretation of ferromagnetism
13. The macroscopic currents' energy in the presence of magnetic materials.

Problems

- 9.1. A small piece of iron is suspended by a wire between the poles of a magnet in a region of uniform field. Is it attracted by the North pole? Or by the South pole? Or by neither of them?
- 9.2. A cube is uniformly magnetized. What is the magnetization at its center?
- 9.3. A small solid cylinder is introduced into a magnetic field, directed along its axis. At a point of the axis, the magnetic field is now smaller than in a vacuum. Is the material of the cylinder para-, dia- or ferro-magnetic?
- 9.4. In which of the following cases does the magnetic susceptibility depend on temperature: diamagnetic, paramagnetic, ferromagnetic?
- 9.5. Can the lines of the \mathbf{H} field be closed under time-independent conditions? Can they radiate from a singular point?
- 9.6. Does the inductance of an iron core solenoid depend on: the number of loops, the iron permeability, the temperature, the current intensity, the section of the core?
- 9.7. A cylindrical iron bar is magnetized perpendicular to the axis. If we cut it down the middle, do we obtain a South and a North pole on the new faces?
- 9.8. The poles of a powerful electromagnet have the shape of truncated cones separated by an air gap in which $B = 1.5$ T. Very imprudently, you get close with a wrench in your hand (never do that!). Suddenly, the wrench is pulled out of your hand by the forces of the field (and might hit you). Where does it stick? On the center of a pole, where the field intensity is greatest, or on the rim, where its gradient is at a maximum?
- 9.9. A plate of parallel faces of a homogeneous material of permeability μ is introduced into a magnetic field \mathbf{B}_0 directed normally to its faces. Determine \mathbf{B} and \mathbf{H} in the plate.
- 9.10. A long and thin iron stick has a magnetization \mathbf{M} parallel to its axis. Express the \mathbf{H} and \mathbf{B} fields inside the stick.
- 9.11. Figure 9.28 represents an iron core (with a gap) around which a number of turns are wound, carrying a steady current. Its section is A , its magnetic permeability is μ . What are the fluxes of \mathbf{B} and \mathbf{H} through the surface S shown in the figure?
- 9.12. A solenoid in air has inductance L . How much does the inductance change if one fills the solenoid with a nucleus of permeability equal to 500?

Fig. 9.28 The system of problem 9.13

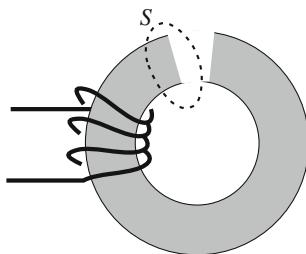
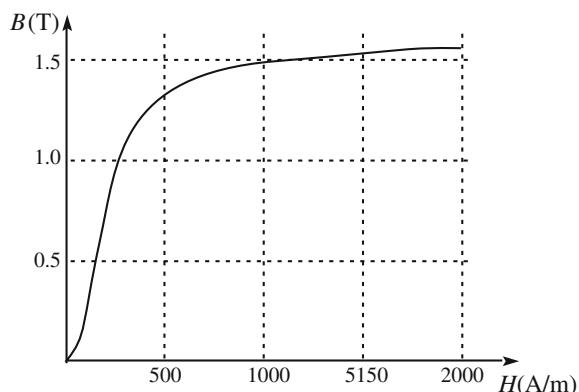


Fig. 9.29 The magnetization curve for problem 9.14



- 9.13. The poles of an iron core electromagnet have a surface of a few square decimeters and a gap width of 1 cm. How much does the \mathbf{B} field in the gap change if we double the width of the gap, while keeping the current intensity constant?
- 9.14. Figure 9.29 shows the magnetization curve for a certain type of iron. How would you determine it experimentally? Starting from the diagram in Fig. 9.28, draw the dependence of the permeability on H . Find the maximum value of the permeability and the corresponding value of H .

Chapter 10

Maxwell Equations

Abstract In this chapter, we discuss the Maxwell equations that completely describe all electric and magnetic phenomena. Maxwell found, in particular, that the equation of the curl of the magnetic field valid under time-independent conditions cannot hold in dynamics, being incompatible with electric charge conservation. Maxwell solved the problem by introducing a new term. This term not only predicts the existence of electromagnetic waves, the waves we use for our radio, television and cellular phones, but also that light is an electromagnetic wave. Electricity, magnetism and optics became unified under a single theory.

In this chapter, we study the Maxwell equations that completely describe all electric and magnetic phenomena and that we have been gradually learning over the course of these lectures. These four partial differential equations give the divergence and the curl of the electric and magnetic field at every point in space and in each time instant. The two equations for the divergences we learnt under time-independent conditions and the Faraday equation for the curl of the electric field are valid without any modification, in general. Contrastingly, we shall find that the equation for the curl of the magnetic field valid under time-independent conditions cannot hold in dynamics. Indeed, it would be incompatible with electric charge conservation, which is a well-established universal law of physics. The problem was solved by Maxwell, who introduced a new term, to which he gave the name of displacement current. The term (which does not correspond to any physical current) has enormous importance. It completely links electric and magnetic fields, which become tightly connected in a unique physical entity, the electromagnetic field. New physical phenomena are foreseen, in particular, the propagation of the fields in an empty space with the speed of light. These are electromagnetic waves, which we use for our radio, television and cellular phones. The Maxwell theory showed, in addition, that light is just a type of electromagnetic wave. Electricity, magnetism and optics became unified under a single theory.

In Sect. 10.2, we shall start with the example of an LC circuit driven by an alternate emf at high frequency and see that it radiates an electromagnetic field propagating throughout the surrounding space. While at low frequencies, electric and magnetic fields are spatially separated and confined, one inside the capacitor, one inside the inductor, at high frequencies, the fields invade the surrounding space,

generating one another. We shall then show how Maxwell equations predicted the existence of electromagnetic waves, which have their speed in a vacuum equal to $1/\sqrt{\epsilon_0\mu_0}$. As such, Maxwell could calculate this value from the known values of the two constants, which had been measured in electrostatic and magnetostatic phenomena, respectively. He found a value quite close to the speed of light. However, the experimental uncertainties in those constants were still sizeable. Maxwell needed a more precise control for his theory. To this aim, he designed an elegant experiment to measure $1/\sqrt{\epsilon_0\mu_0}$ directly, which we describe in Sect. 10.3.

In Sect. 10.4, we find the general expressions of the electromagnetic field energy density and flux. We shall see, in particular, how the field energy flows not only under dynamic conditions but also when nothing depends on time. We discuss some simple examples of that in Sect. 10.5. Similarly, in Sect. 10.6, we see that the electromagnetic field also stores linear momentum and that the linear momentum density is proportional to the energy flux. As a consequence, the energy, linear and angular momentum conservation laws for an isolated system of charged particles do not hold if we only include the mechanical (namely that of the particles) energy, linear and angular momentum, respectively. This is because each of these quantities can be exchanged between matter (particles) and field during the evolution of the system. However, the conservation laws rigorously hold when we consider the total energy, namely the sum of the particles and field energy, the total linear momentum of the particles and field, and the total angular momentum of the particles and field.

In Sect. 10.7, we re-express the Maxwell equations in a form that is useful in the presence of matter, in terms of the macroscopic charge and current densities and of the auxiliary fields **D** and **H**. In Sect. 10.8, we study the discontinuities of the electric and magnetic fields in crossing a charged surface and a current-carrying surface, respectively.

In Sect. 10.9, we find the general expressions of the scalar and vector potentials and the differential equations ruling them. We shall find, once more, how the equations for the potentials, which are equivalent to the Maxwell equation for the fields, predict electromagnetic waves propagating in a vacuum with the speed of light. In Sect. 10.10, we discuss the physical meaning of the potentials and find that they are, respectively, the potential interaction energy and the potential interaction momentum of the unit charge.

The relativity principle holds for the Maxwell equations in the form originally established by Galilei only if the transformations of space coordinates and time between two inertial frames are the Lorentz rather than the Galilei transformations. Indeed, this was the historical reason that led H. Lorentz to his fundamental discovery. As a matter of fact, the equations describing electromagnetic phenomena were historically developed in a form already valid at high velocities and had no need to be modified by relativity, as opposed to the case of the mechanical laws (see 1st volume of the course). We shall show the Lorentz covariance properties of electromagnetism on the equations for potentials, rather than on those for fields, because the former is more immediate and easier.

10.1 Displacement Current

In this chapter, we shall complete the set of equations that describe the electromagnetic field in full generality. These partial differential equations ruling the electric and magnetic fields are the famous Maxwell equations. James Clerk Maxwell (Scotland UK, 1831–1879) worked on their development for 10 years between 1855 and 1865, when he published the complete theory in the paper “On the dynamical theory of electromagnetic field”.

As a matter of fact, we already know the Maxwell equations almost completely, but a fundamental piece of them is still missing. We learnt two equations for the divergence and the curl of the electric field, namely

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \quad (10.1)$$

and

$$\nabla \times \mathbf{E} = - \frac{\partial \mathbf{B}}{\partial t} \quad (10.2)$$

and two equations for the divergence and the curl of the magnetic field, namely

$$\nabla \cdot \mathbf{B} = 0. \quad (10.3)$$

and

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j}. \quad (10.4)$$

We have established, in particular, Eq. (10.4) under time-independent conditions. If valid under dynamic conditions as well, the equation would imply that the electric charge is not conserved, in contradiction to the experimental evidence. It was not easy to show that in Maxwell’s time, but we can do it now very simply. Let us take the divergence of the two sides of Eq. (10.4) and let us remember that the divergence of a curl is identically zero. We immediately obtain that $\nabla \cdot \mathbf{j} = 0$. This means that the current density is solenoidal everywhere, or, in other words, that the current lines are always and only closed lines. But this is in contradiction with the (local) charge conservation, which is expressed by the continuity equation, namely

$$\nabla \cdot \mathbf{j} = - \frac{\partial \rho}{\partial t}. \quad (10.5)$$

Hence, the divergence of \mathbf{j} is zero only under steady conditions. Maxwell discovered the problem and proposed the solution. The logical process he followed to this fundamental discovery was not, however, as simple as it looks today. It is worthwhile analyzing it briefly. Note that, at the time, “understanding” a physical phenomenon meant having a mechanical model capable of explaining it. Maxwell

developed a mechanical model in which the electromagnetic field was similar to a complex elastic medium, composed of substructures capable of polarizing under the action of an \mathbf{E} field. These substructures are not the molecules we know. The polarization is a consequence of a displacement of the positive charges relative to the negative ones. For this reason, \mathbf{D} is called electric displacement. If \mathbf{E} varies with time, this relative displacement of the charges varies in time as well. The motion of the charges is a microscopic current, which Maxwell called the *displacement current*. Once the theory was fully developed, the mechanical model was abandoned in the same way as scaffoldings are dismantled once the cathedral has been built. The names “electric displacement” and “displacement current” remained, even if we know that these quantities have nothing to do with physical displacements or with physical currents.

For us, it is now easy to see that the difficulty disappears by adding the term $\mu_0\epsilon_0\partial\mathbf{E}/\partial t$ to the right hand side of Eq. (10.4). We obtain

$$\nabla \times \mathbf{B} = \mu_0\mathbf{j} + \mu_0\epsilon_0 \frac{\partial \mathbf{E}}{\partial t}. \quad (10.6)$$

Let us check by again taking the divergence of both sides of the equation. We get

$$0 = \nabla \cdot \mathbf{j} + \epsilon_0 \nabla \cdot \frac{\partial \mathbf{E}}{\partial t}.$$

We invert the time derivative with the divergence in the last term and use Eq. (10.1), obtaining

$$0 = \nabla \cdot \mathbf{j} + \epsilon_0 \frac{\partial \nabla \cdot \mathbf{E}}{\partial t} = \nabla \cdot \mathbf{j} + \frac{\partial \rho}{\partial t},$$

which is just the charge conservation of Eq. (10.5). The introduction of the new term, which was originally based on theoretical arguments, has enormous consequences. It implies the existence of new classes of phenomena, which were later experimentally observed, confirming the correctness of the theory. We shall see several of them below.

Equations (10.1), (10.2), (10.3) and (10.6), namely the Maxwell equations, describe in complete generality the electric and magnetic fields. They are the fundamental equations of electromagnetic interaction. The equations are, in principle, capable of quantitatively describing all electromagnetic phenomena at the classical level, electric charge conservation included.

We already know the integral form of the first three Maxwell equations. The fourth one has an integral expression as well. Let Γ be an arbitrarily oriented closed curve and Σ a surface bound by Γ , oriented with the positive direction in agreement with that of Γ . Let us calculate the flux through Σ of both sides of Eq. (10.6).

As usual, let us transform the flux of the curl of \mathbf{B} through Σ in the circulation of \mathbf{B} about Γ . We obtain

$$\oint_{\Gamma} \mathbf{B} \cdot d\mathbf{s} = \mu_0 I_{\text{linked}} + \mu_0 \epsilon_0 \frac{\partial \Phi_E}{\partial t}, \quad (10.7)$$

which we read as: the circulation of \mathbf{B} about any closed curve Γ is equal to μ_0 times the current intensity linked to Γ plus $\epsilon_0 \mu_0$ times the rate of change of the flux of \mathbf{E} through Γ . The law generalizes the Ampère law in dynamic situations. The term $\epsilon_0 d\Phi_E/dt$ clearly has the physical dimensions of a current intensity. It is called the displacement current, even if it has nothing to do with either currents or displacements.

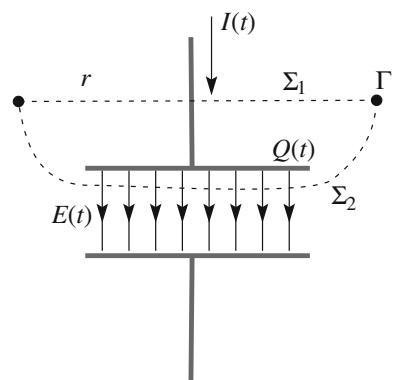
Let us now look at an example as to how the displacement current is able to get rid of a problem that would be present without it. Consider a parallel plate capacitor of circular plates. The capacitor is being charged by a current, whose intensity $I(t)$ varies with time. The current is carried by straight wires perpendicular to the plates, as in Fig. 10.1.

Let us start by assuming the Ampère law to hold without the displacement current term. Let us apply the law to the curve Γ , which is an oriented circle perpendicular to and centered on the wire, as shown in the figure. The law must be valid for any surface having Γ as a boundary. Let us first calculate the fluxes of both sides of Eq. (10.4), choosing the circle Σ_1 in the plane of Γ . It is crossed by the current I . We obtain

$$\int_{\Sigma_1} \nabla \times \mathbf{B} \cdot \mathbf{n} d\Sigma = \oint_{\Gamma} \mathbf{B} \cdot d\mathbf{s} = \mu_0 I.$$

Let us now try the same with the surface Σ_2 , which has Γ as a boundary as well, but does not intercept the current, going through the gap between the plates. We now obviously have

Fig. 10.1 The circle Γ cuts the page at the two dots. Σ_1 and Σ_2 are two surfaces circled by Γ



$$\int_{\Sigma_2} \nabla \times \mathbf{B} \cdot \mathbf{n} d\Sigma = \oint_{\Gamma} \mathbf{B} \cdot d\mathbf{s} = 0.$$

The contradictory result stems from the fact that \mathbf{j} is not solenoidal. Indeed, its lines end on and originate from the plates of the capacitor. Let us now also consider the new term, as put forth in Eq. (10.6). Nothing changes with Σ_1 , because on that surface, the field \mathbf{E} and the rate of change of its flux are zero. Through Σ_2 , the current is still zero, but the rate of change of the flux of the electric field \mathbf{E} is not so inside the gap. Let us calculate it, neglecting the fringe effects. The field intensity between the plates is $E = \sigma/\epsilon_0$, where the surface charge density on the plates is $\sigma = Q/A$, where A is the area of the plates and Q the charge. The flux of the electric field is then $\Phi_E = EA = (\sigma/\epsilon_0) A = Q/\epsilon_0$ and we obtain

$$\int_{\Sigma_2} \nabla \times \mathbf{B} \cdot \mathbf{n} d\Sigma = \oint_{\Gamma} \mathbf{B} \cdot d\mathbf{s} = \mu_0 \epsilon_0 \frac{d\Phi_E}{dt} = \mu_0 \frac{dQ}{dt} = \mu_0 I.$$

Note that in the final step, we have assumed charge conservation, which states that the rate of change of the charge of the plate is equal to the charge reaching the plate in a second, which is the current intensity, namely that $dQ/dt = I$.

In conclusion, the displacement current inside the capacitor takes the place of the conduction current in the wires, solving the problem we started with. We note that the lines of \mathbf{B} above and below the capacitor are circles normal to and centered on the conduction current. The lines in the plane of the capacitor are of the same type, encircling the rate of change of the flux of \mathbf{E} in place of the conduction current.

QUESTION Q 10.1. The voltage between the plates of a parallel plate capacitor in a vacuum varies at a rate of 100 V/s. The area of the plates is 10 cm^2 and their distance is 1 mm. How much is the displacement current inside the capacitor? □

Figure 10.2 shows a schematic idea for a possible experiment to verify Eq. (10.6). A toroidal solenoid is in the middle plane between the circular plates of a parallel plate capacitor (position 1 in the figure). The capacitor is connected to an alternate (sinusoidal) emf source. Consequently, the electric field between the plates varies with time, producing a magnetic field according to Eq. (10.6). The lines of the magnetic field are circles around the axis of the system, as just discussed. Some of them are inside the solenoid. The corresponding flux of \mathbf{B} linked to the solenoid varies with time as well, inducing an emf, according to the Faraday law.

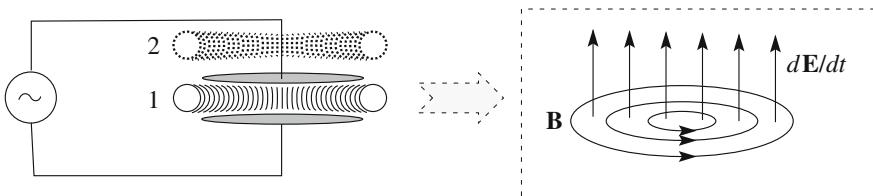


Fig. 10.2 Schematic idea for detecting displacement current

We measure the induced emf under these conditions. We then move the solenoid upward out of the capacitor (position 2 in the figure) and measure the emf induced by the variable current. The electromotive forces are found to be equal to one another.

Note that the effect to be measured increases with the frequency of the alternate current. In practice, the effect is important only at quite high frequency and is usually negligible at network frequency.

QUESTION Q 10.2. Do you expect to measure any effect if you apply an emf increasing with time at a constant rate in place of an alternate one to the capacitor shown in Fig. 10.2? Why? \square

10.2 Electromagnetic Waves

The enormous consequences of the new term appear when one considers the complete set of the Maxwell equations. Let us start by looking closely at the two universal constants they contain, the vacuum permittivity ϵ_0 and the vacuum permeability μ_0 . We encountered the former in electrostatics. The force between two point-charges in a vacuum is inversely proportional to ϵ_0 (or directly proportional to $1/4\pi\epsilon_0$). We encountered the latter in magnetostatics. The force between two parallel straight current-carrying wires is directly proportional to μ_0 . Electric and magnetic fields that were independent entities under steady conditions become a unique physical system linked by Eqs. (10.2) and (10.6) under dynamic conditions.

The new term, the displacement current, is proportional to $\epsilon_0\mu_0$, namely to both the electric and the magnetic constants. Let us analyze the physical dimensions of this product. From Eq. (1.8), we see that ϵ_0 has the dimensions of the inverse of a force times an inverse square length times a square charge. We can then write the dimensional equation in terms of the SI base units (mass, length, time and current intensity) as

$$[\epsilon_0] = [M^{-1}L^{-3}T^4I^2]. \quad (10.8)$$

From Eq. (6.23), we see that μ_0 has the dimensions of a force divided by a current intensity squared. We can then write

$$[\mu_0] = [MLT^{-2}I^{-2}]. \quad (10.9)$$

The dimensions of the product of the two constants is then

$$[\epsilon_0\mu_0] = [L^{-2}T^2]. \quad (10.10)$$

The electric units have disappeared into the product. The product is dimensionally an inverse square velocity. Calculating this velocity with the values of Eqs. (1.9) and (6.23), we obtain

$$c = 1/\sqrt{\epsilon_0 \mu_0} = 2.998 \times 10^8 \text{ ms}^{-1}. \quad (10.11)$$

Indeed, this is an extremely important velocity, one of the fundamental constants of physics, the velocity of light. We shall come back to that shortly.

Let us now analyze the other possible combination of the two constants, which is their ratio. This ratio is, in fact, proportional to the product of the constants in front of the Coulomb force between charges and the Ampère law between currents. We have

$$[\mu_0/\epsilon_0] = [M^2 L^4 T^{-6} I^{-4}]. \quad (10.12)$$

The combination on the right-hand side looks quite complicated, but, in fact, it is not. These are the physical dimensions of an electric resistance square. Indeed, the resistance is an electric potential, namely an electric field times a distance, divided by a current intensity. The electric field is a force divided by a charge, and we can write

$$[R] = [FQ^{-1} LI^{-1}] = [MLT^{-2} I^{-1} T^{-1} LI^{-1}] = [ML^2 T^{-3} I^{-2}]. \quad (10.13)$$

Let us calculate the resistance of which Eq. (10.11) is the square. We find

$$Z_0 = \sqrt{\mu_0/\epsilon_0} = 376.73 \text{ k}\Omega. \quad (10.14)$$

This constant is called the *impedance of the free space* or *vacuum impedance*. As we shall study in the 4th volume of the course, this quantity relates the magnitudes of the electric and magnetic fields of electromagnetic waves propagating in a vacuum. In the next section, we shall see how Maxwell experimentally confirmed his theory of electromagnetic waves by measuring the vacuum impedance under purely static conditions.

Let us now go back to the Maxwell equations and start our discussion of the consequences of the new term. Let us consider a point-charge q at rest in the origin of the reference system. The electric field \mathbf{E} in the entire space is the well-known electric field of a point-charge, while the magnetic field \mathbf{B} is zero everywhere. Suppose that the charge starts moving at a certain moment. In the regions close to the charge, the electric field now varies with time. Its rate of change generates a curl of magnetic field according to the Faraday law (Eq. 10.2), and hence a non-zero magnetic field \mathbf{B} . In general, \mathbf{B} will vary with time as well, giving origin with its rate of change by Eq. (10.6) to a new electric field \mathbf{E} . Notice that this contribution to the electric field does not have the charge as a source. This electric field is not constant, but varies with time, giving origin to a \mathbf{B} , which varies with time as well, giving origin to an \mathbf{E} , and so on and so on. This chain of processes produces an electromagnetic field (namely an electric and a magnetic field that are intimately linked), which varies with time and propagates in space over greater and greater distances from the point-charge that gave origin to the phenomenon. This is

analogous to when you toss a stone into the water of a lake. Soon, a few circular waves appear around the point at which the stone hit the water. The waves then move out, increasing their radii, with circles that spread out along the surface of the lake. The propagation continues well beyond the point at which the stone has ceased to move, until the borders of the lake absorb the energy transported by the waves. As we shall see at the end of this section, the electromagnetic waves in a vacuum have a well-defined velocity, given by Eq. (10.11).

Let us analyze an example useful for understanding the process qualitatively. Consider the circuit in Fig. 10.3 consisting of a solenoid (inductance L), a capacitor (capacitance C) and an alternate (sinusoidal) emf (\mathcal{E}) regenerator at the angular frequency ω .

Note that inductance and capacitance are physical quantities, which are well defined when the regime is quasi-steady, namely when the rate of change of \mathcal{E} , and consequently of the current intensity I , are small enough to be equal, at any given instant, in all the sections of the circuit. Let us then assume ω to be small. The electric field under these conditions is not very different from the electrostatic field of the capacitor charged with the instantaneous value of the charge on its plates. The field is substantially contained inside the capacitor. Similarly, the magnetic field is substantially the field of a steady current and is contained inside the solenoid.

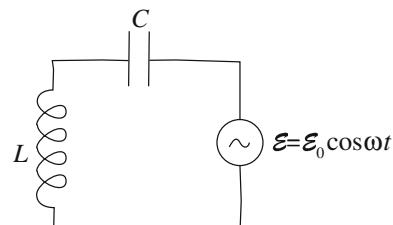
If we now increase the frequency ω , namely if we increase the rate of change of the emf, the charge on the plates of the capacitor, and the field \mathbf{E} they produce, will vary more quickly, as will the rate of change of the current in the solenoid and the field \mathbf{B} (Fig. 10.4).

The electric field in the capacitor varying in time produces lines of \mathbf{B} , running about the capacitor and invading the external space. The varying magnetic field \mathbf{B} in the solenoid produces lines of \mathbf{E} wrapping the solenoid in the external space. In this way, the fields start to invade the space external to the circuit elements. The circuit radiates electromagnetic waves.

One can increase the radiation by opening the capacitor, moving its plates away from one another. In doing so, the electric field will occupy a larger region of space. Similarly, one can substitute the solenoid with a straight conductor, which produces a less localized magnetic field. This is shown in Fig. 10.5.

Even better, one can eliminate the capacitor completely and connect directly to the poles of the generator with metallic bars, as in Fig. 10.6.

Fig. 10.3 A LC circuit in alternated current



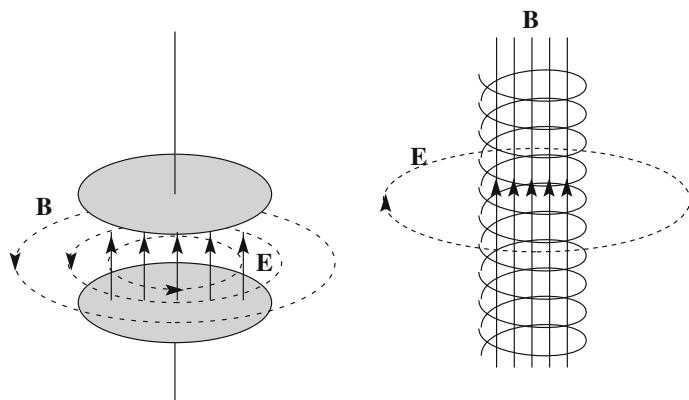


Fig. 10.4 Magnetic field lines around the capacitor and magnetic field lines around the solenoid

Fig. 10.5 Opening up the capacitor

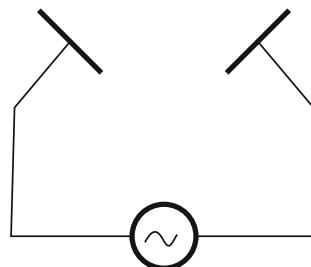


Fig. 10.6 An electromagnetic aerial



Notice that, at the moment in which the bar on the left of the generator is positively charged, the one on the right is negative, and vice versa. We have made an electromagnetic aerial. Maxwell equations tell us that electric fields are generated by electric charges and magnetic fields varying with time, and that magnetic fields are generated by currents and electric fields varying in time. At a low enough frequency, the electric and magnetic fields are due mainly to the charges and the currents. At higher frequencies, the fields near the aerial are still dominated by charges and currents, but in the far away space, the electric field is dominated by $\partial\mathbf{B}/\partial t$ and the magnetic field by $\partial\mathbf{E}/\partial t$. The fields born in this way invade wider and wider regions of space over time, being generated by one another. Once it has been produced by its source, the electromagnetic field takes on, if we can put it this way, a life of its own, and will continue to propagate infinitely even if the generator is shut off or the aerial is destroyed.

Let us now analyze how the Maxwell equations quantitatively describe the sort of interplay between electric and magnetic fields of which we just gave an example. This is the case in which the two fields generate each other, even in a space where charges and currents are absent, namely in a vacuum. Under these conditions, the Maxwell equations (10.1), (10.2), (10.3) and (10.6) become

$$\nabla \cdot \mathbf{E} = 0, \quad (10.15)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad (10.16)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (10.17)$$

and

$$\nabla \times \mathbf{B} = \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}. \quad (10.18)$$

Let us take the curl of both sides of Eq. (10.16), obtaining

$$\nabla \times \nabla \times \mathbf{E} = -\frac{\partial(\nabla \times \mathbf{B})}{\partial t}.$$

We now use the vector identity

$$\nabla \times \nabla \times \mathbf{E} = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E}$$

and express $\nabla \times \mathbf{B}$ on the right-hand side with Eq. (10.18). We also use Eq. (10.15) for $\nabla \cdot \mathbf{E}$, obtaining

$$\frac{\partial^2 \mathbf{E}}{\partial t^2} - \frac{1}{\mu_0 \epsilon_0} \nabla^2 \mathbf{E} = 0. \quad (10.19)$$

A similar argument, starting from Eq. (10.18), leads to the same equation for the magnetic field, namely to

$$\frac{\partial^2 \mathbf{B}}{\partial t^2} - \frac{1}{\mu_0 \epsilon_0} \nabla^2 \mathbf{B} = 0, \quad (10.20)$$

which is the same equation. This partial differential equation is famous, known as wave equation. We cannot discuss its solution here without a deeper knowledge of the physics of waves, which is the subject of the 4th volume of this course. Here, we will simply state that the equation foresees the existence of electromagnetic waves, propagating in a vacuum with speed

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}. \quad (10.21)$$

The two constants on the right-hand side are measured with electrostatic and magnetostatic experiments, respectively. Knowing these values, Maxwell not only predicted electromagnetic waves, but also found that their velocity should have been exactly equal, within experimental errors, to the speed of light. Maxwell concluded that light must be an electromagnetic phenomenon, unifying electromagnetism and optics, two branches of physics that had been completely separate before him.

However, in 1865, the experimental values of both sides of Eq. (10.21) were known with rather limited accuracy. Maxwell was not only one of the greatest theoretical geniuses, but a ingenious experimenter as well. We have already mentioned, in Sect. 5.4, that Maxwell had been appointed to the British Committee for electrical standards in 1862 and had been working under William Thomson on the British Association Ohm standard for electric resistance. Having developed his theory, Maxwell turned to trying to improve its experimental check. We shall discuss his work in the next section.

10.3 The Maxwell Ratio of Units Experiment

This section, which has a historical charter, will deal with the experimental control by Maxwell of Eq. (10.21).

The left-hand side of Eq. (10.21) is the velocity of light in a vacuum. In the 1860s, two laboratory measurements had been done. In 1849, Hippolyte Fizeau (France, 1819–1996), as we shall describe in the 4th volume, had measured the value $c = 3.15 \times 10^8$ m/s. In 1862, in a more accurate experiment, Léon Foucault (France, 1819–1868) had obtained $c = 2.98 \times 10^8$ m/s. The two results differ by about 5 %.

The quantity on the right-hand side had been determined by Rudolf Kohlrausch (Germany, 1809–1858) and Wilhelm Weber (Germany, 1804–1891) in 1856. Those authors had measured the potential difference of a capacitor of known capacitance, thereby establishing the charge electrostatically. The capacitor was then discharged through a ballistic galvanometer, measuring the same charge as current intensity integrated over time. The result was $(\epsilon_0 \mu_0)^{-1/2} = 3.11 \times 10^8$ m/s. As Maxwell puts it, “the only use made of light in the experiment was to see the instruments”.

While the values were not in disagreement, their experimental uncertainties were large, and Maxwell thought he had to improve the situation. The result was an ingenious and elegant experiment. The outcome was presented to the Royal Society of London in 1868 under the title, “On a direct comparison of electrostatic with electromagnetic force”.

Maxwell’s style in designing experiments was defined by the use of a null method. We saw an example of that in Sect. 6.6. In the null method, the effects of two processes counterbalance one another, leading to a null result if they are exactly

equal. The method can be very sensitive, because the smallest difference between the two effects shows up as an unbalance. In this experiment, the null method is employed in two different ways.

Let us read and explain the relevant parts of the paper.

The basic idea of the apparatus was the following.

The experiments consisted in observing the equilibrium of two forces, one of which was the attraction between two disks, kept at a certain difference of potential, and the other was the repulsion between two circular coils, through which a certain current passed in opposite directions.

A basic design principle was to avoid absolute measurements and to rely only on relative ones. Let us be more precise. The electrostatic force, including the force between the two charged disks, is proportional to $1/\epsilon_0$. The magnetostatic force between two steady currents, including the repulsion between the coils of the experiment, is proportional to μ_0 . Equation (10.11) told us that the ratio of the two proportionality constants, namely $1/(\epsilon_0 \mu_0)$, the quantity we want to determine, has the dimension of the square of a velocity, while Eq. (10.14) tells us that their product, namely μ_0/ϵ_0 , has the dimension of a resistance square. The Maxwell experiment measures the value of this resistance relative to a standard, the British association ohm.

We mention here that in the period of development of electromagnetism, when electricity and magnetism were still two separate branches, two different units for the electric charge had been defined. The “electrostatic unit” (esu) had been defined on the basis of the Coulomb law, the “electromagnetic unit” (emu) on the basis of the Ampère law on the force between two current-carrying wires. The ratio between those units resulted in being just $1/(\epsilon_0 \mu_0)^{1/2}$. Consequently, the Maxwell experiment we are describing and similar ones are called ratio of units experiments.

Let us read what Maxwell writes:

In the experiments here described no absolute measurements were made, either of length, time, or mass, the ratios only of these quantities being involved; and the velocity determined is expressed in terms of the British Association Unit of resistance, so that whatever corrections may be discovered to be applicable to the absolute value of that unit must be also applied to the velocity here determined.

Let us see how that can be done. The electrostatic force in the Maxwell experiment is the attractive force between the plates of a parallel plate capacitor. Let Q be its charge, V the corresponding potential difference, C the capacitance, r the radius of the plates, and h their distance. To be precise, Maxwell used the trick of the guard ring, in order to avoid any appreciable fringe effect. Consequently, the area to be considered is the area of the plate inside the ring. Equation (3.13) gives the (electric) force on that plate as $F_e = Q^2/(2\epsilon_0 \pi r^2)$, and Eq. (2.18) the capacitance $C = \epsilon_0 \pi r^2/h$. The force is then

$$F_e = \frac{\pi r^2}{2 h^2} \epsilon_0 V^2 = k_1 \epsilon_0 V^2, \quad (10.22)$$

where we have defined the constant $k_1 = (\pi/2)(r^2/h^2)$, which is a dimensionless quantity. It depends on the ratio of two lengths, the radius of the plates over their distance.

The magnetic force is the repulsive force between two circular coils carrying the same current in opposite directions. Let l be the total length of the coils, d their distance and I the current intensity. We know the force between two steady currents for a different geometry, namely two straight parallel currents. Equation (6.56) tells us that the force on a length l of the wires is $F_m = \frac{l}{2\pi d} \mu_0 I^2$. Once again, the constant $l/(2\pi d)$ is dimensionless. The magnetic force between two circular coils is similar. The difference is in the geometrical factor, which is not $l/(2\pi d)$, but a different dimensionless function of l/d . The calculation of this function involves a quite difficult integral. This is an elliptic integral (as it is called), which cannot be evaluated analytically. Maxwell calculated numerically up to the required accuracy. Let us call k_2 this dimensionless constant and write

$$F_m = k_2 \mu_0 I^2. \quad (10.23)$$

Herein lies Maxwell's first brilliant idea, namely to have the potential difference V between the plates of the capacitor in Eq. (10.22) generated by the voltage drop equal to the current I feeding the coils in Eq. (10.23) through a standard resistor R , calibrated in terms of the British Association ohm. We can then substitute $V = RI$ in Eq. (10.22), which becomes

$$F_e = k_1 \varepsilon_0 I^2 R^2. \quad (10.24)$$

The “null” experiment is designed to have the two forces balancing one another. When this condition is reached, it is $k_2 \mu_0 I^2 = k_1 \varepsilon_0 I^2 R^2$. The current intensity can be simplified off. Calling $k_3 = k_1/k_2$, which is another dimensionless constant, we finally have

$$\mu_0 / \varepsilon_0 = k_3 R^2. \quad (10.25)$$

Let us look at the apparatus now. The plates of the capacitor were two circular disks, one fixed to a standing support, one to an arm of a torsion-balance. The two coils were just behind the two disks (electrically insulated from them), one on the fixed support, one on the arm of the torsion-balance. The equilibrium between the two forces could be obtained establishing the right distance between the fixed plate and the plate on the balance arm. This delicate operation (there was a potential difference of more than a kilovolt between the plates) was done with a calibrated micrometer. The micrometer reading at equilibrium was one of the two pieces of data resulting from each experiment. The distances between the plates and the coils were calculated from this datum and the fixed dimensions known by construction. As already mentioned, the capacitor had a guard ring. This “trick” was used by Thomson to build his electrometers, a fact that Maxwell quotes.

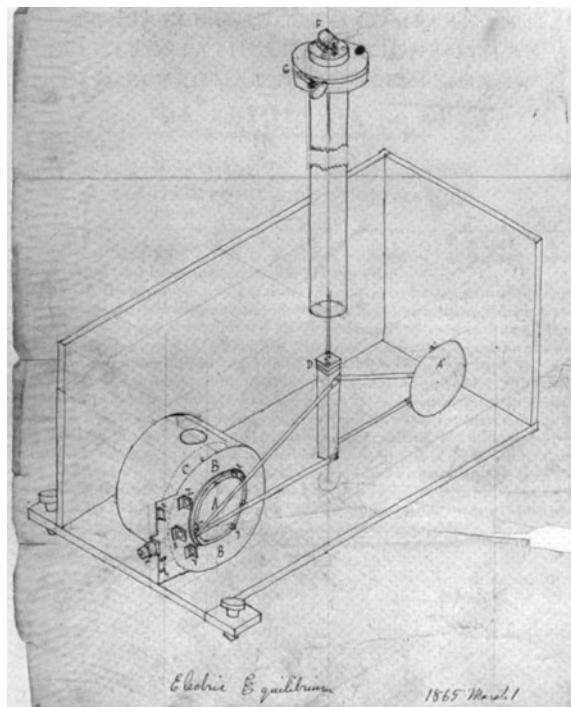
But there is another effect to cancel. When the coil carries current, it is a magnetic dipole. The earth's magnetic field, in which it is immersed, exerts a torque on the coil, which must be canceled. To do that, Maxwell suspended an identical coil from the other extreme of the bar of the torsion-balance, at the same distance from the suspension wire. He fed it with the same current in the right direction so as to have an equal and opposite torque. A mechanical counterweight was also included. Let us read what he wrote about this.

For this purpose, one of the disks, with one of the coils attached to his hinder surface, was suspended on one arm of a torsion-balance, while the other disk, with the other coil behind it, was placed at a certain distance, which was measured with a micrometer-screw. The suspended disk, which was smaller than the fixed disk, was adjusted so that in its position of equilibrium its surface was in the same plane with that of a 'guard-ring', as Sir. W. Thomson's electrometers, and its position was observed by means of a microscope directed on a graduate glass scale attached to the disk. In this way its position could be adjusted to thousandths of an inch, while a motion of much smaller extent was easily detected.

An exactly similar coil was placed at the other end of the torsion-balance, so as to get rid of the effects of the terrestrial magnetism.

A sketch of the mechanical components of the apparatus is shown in Fig. 10.7

Fig. 10.7 The mechanics of the apparatus of the Maxwell experiment. The cylindrical structure on the left end fixed to the floor includes the fixed plate and coil. The mobile plate attached to the torsion-bar is visible with its guard ring. The coil compensating for the earth field effect and the counterweight are on the right end of the torsion-balance frame. The case in which the apparatus was closed is shown cut away

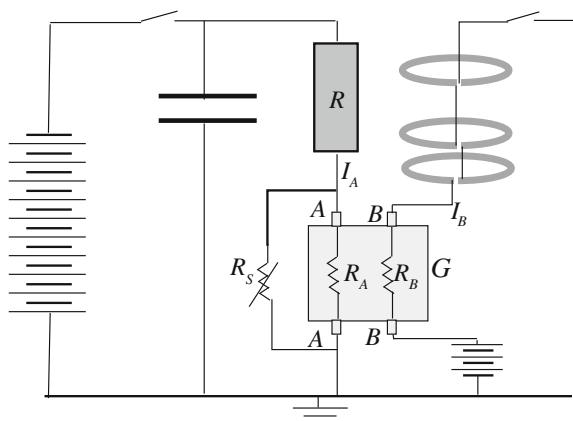


The main electric components of the apparatus were two batteries, a large one to provide the high voltage between the disks and a small one to provide the current to the coils, a calibrated resistor R and a galvanometer.

The galvanometer was based on the same principle we discussed in Chap. 6, shown in Fig. 6.13, but with the important difference that it had two, rather than one, coils carrying two different currents. Maxwell was again using the null method. The two coils of the galvanometer are mechanically joined to one another and their currents generate two opposite torques, which cancel one another out when the two intensities are in a ratio defined by the ratio between the (known) number of turns of the two coils. Figure 10.8 schematically shows how the galvanometer (G) is operated. The two currents, say I_A and I_B , are injected through the terminals AA and BB, respectively. R_A and R_B are the resistances of the two galvanometer coils, which were much smaller than R and had been accurately measured. The variable resistor R_S , called a shunt, is connected, in parallel to R_A , to the AA terminals. Consequently, I_A is divided between an internal part, which produces the torque, and an external part that is mechanically inactive. The smaller battery produces the current I_B that feeds the big coils determining the repulsive force. The equilibrium between I_B and the internal fraction of I_A is reached by adjusting R_S . The value of the shunt resistance at equilibrium is the second datum of each experiment.

Figure 10.8 shows the two circuits. Maxwell used two batteries. Particularly demanding, for the time, was the generation of the voltage of a few kilovolts needed to produce a force of sufficient strength between the disks. The biggest battery in England was owned by John Peter Gassiot, a London wine merchant, who dedicated time and resources to science. Mr. Gassiot was happy to host the experiment in his private laboratory, furnishing Maxwell with a big battery of 2600 voltaic cells of bichloride of mercury, giving up to 3 kV. A smaller battery (9 cells) was used to produce the current for the coils. The calibrated reference resistor (a coil of wire) was provided by Mr. Willoughby Smith. The currents of the two circuits entered the two inputs of the galvanometer.

Fig. 10.8 The electrical arrangement of the Maxwell ratio of units apparatus. The disks of the condenser, the three coils and the scheme of the galvanometer G are shown



J. Clerk Maxwell writes

The electrical arrangement was the following.

One electrode of Mr. Gassiot's great battery was connected with a key (*meaning switch*). When the key was pressed connection was made to the fixed disk, and thence, through Mr. Willoughby Smith's resistance-coils, to a point where the current was divided between the principal coil of the galvanometer and a shunt, S , consisting of Mr. Jenkin's resistance coils (*the calibrated resistor*). These partial currents reunited at a point where they were put in connection with the other electrode of the battery, ..., and with earth.

Another battery was employed to send a current through the coils (*the two big ones*). One electrode of the battery was connected with a second contact piece of the key, so that, when the key was pressed, the current went first through the secondary coil of the galvanometer, consisting of thirty windings of thick wire, then through the fixed coil, ...and so through the two suspended coils to the case, to earth, and to the other electrode of the battery.

When these arrangements had been made, the observer at the microscope, when the suspended disk was stationary at zero, made simultaneous contact with both batteries by means of the key. If the disk was attracted, the great battery was the more powerful, and the micrometer was worked so as to increase the distance of the disk. If the disk was repelled, the fixed disk had to be moved nearer to the suspended disk, till a distance was found at which, when the scale was at rest and at zero, no effect was produced by simultaneous action of the batteries.

In the meantime the other observer at the galvanometer was taking advantage of these contacts to alter the shunt S , till the effect of the two currents on the galvanometer-needle balanced each other.

When a satisfactory case of equilibrium has been observed simultaneously at the galvanometer and at the torsion-balance, the micrometer-reading and the resistance of the shunt were set down as the result of the experiment.

Maxwell then discusses both the difficulties he had and the experimental controls and calibrations done. The calculation of the dimensionless constant that we called k_3 follows, based on the distance measured with the micrometer, the geometrical dimensions known by construction and the resistances of the circuit, different from R , which had been measured.

In this expression the only quantities which must be determined in absolute measure are the resistances. The other quantities which must be measured are the ratios of the radius of the disk to its distance from the fixed disk, and the ratio of the radius of the coils to the distance between them.

Maxwell made 10 experiments under satisfactory conditions. The average of the experimental values of the ratio of units was

$$\frac{1}{4\pi} \sqrt{\frac{\mu_0}{\epsilon_0}} = 28.798 \Omega, \text{ or B.A. units.} \quad (10.26)$$

corresponding to

$$\sqrt{\frac{1}{\mu_0 \epsilon_0}} = 2.88 \times 10^8 \text{ m/s.} \quad (10.27)$$

In conclusion, Maxwell observed that the result is

decidedly less than any estimate of the velocity of light, of which the lowest, that of M. Foucault, is 298,000,000 m/s.

Indeed, precise measurements of both the ratio of units and the speed of light are quite difficult. After the publication of the Maxwell theory, work to increase the accuracy of both started worldwide. In 1878, in the third edition of his “A treatise on electricity and magnetism”, James Clerk Maxwell wrote

It is manifest that the velocity of light and the ratio of the units are quantities of the same order of magnitude. Neither of them can be said to be determined as yet with such degree of accuracy as to enable us to assert that the one is greater than the other. It is to be hoped that, by further experiments, the relation between the magnitudes of the two quantities may be more accurately determined.

In the mean time our theory, which asserts that these two quantities are equal, and assigns a physical reason for this equality, is not contradicted by the comparison of these results such as they are.

By the next year, the year of the Maxwell’s death, the equality of the two quantities had been established with 1 % accuracy. William Ayrton (UK, 1847–1908) and John Perry (UK, 1850–1920) had measured the ratio of units as $1/\sqrt{\mu_0 \epsilon_0} = 2.96 \pm 0.03 \times 10^8 \text{ m/s}$ and Albert Michelson (USA, 1852–1931) the velocity of light (in air) as $c = 2.998\ 64 \pm 0.000\ 51$.

QUESTION Q 10.3. In the Maxwell experiment, the diameter of the plate inside the guard ring was 105 mm and the distance between the fixed and mobile plates 1.65 mm. How much was the force when a voltage of 3 kV was applied? □

A note on the measurement units. Let us briefly come back to the measurements units that we have been using. In principle, the measurement units of the physical quantities are arbitrary. In practice, they are far from being so. For every practical purpose, it is mandatory to have internationally agreed-upon choices. As we mentioned in the 1st volume of this course, international organizations were created to foster international standardization worldwide. The International Conference of Weights and Measures, CGPM, for short, is the body responsible for the definition of the units. As we know, the unit system is called the *Système International* in French, or SI, for short.

We repeat here what we have already stated. In the SI, the values of two basic constants are given by definition. These are the vacuum permeability, which is defined as

$$\mu_0 = 4\pi \times 10^{-7} \text{ N A}^{-2}, \quad (10.28)$$

which is, in fact, the definition of the ampere, and the light velocity in a vacuum, defined as

$$c = 299,792,458 \text{ ms}^{-1}. \quad (10.29)$$

The value of the vacuum permittivity stems from these two definitions as

$$\epsilon_0 = \left[4\pi \times 10^{-7} \times (299,792,458)^2 \right]^{-1} = 8.845187817 \dots \text{ pF m}^{-1}. \quad (10.30)$$

The values of three constants are “exact”, namely they have no experimental uncertainty (to be pedantic, ϵ_0 has the uncertainty stemming from the finite number of figures for which π is known). As a matter of fact, the experimental uncertainties, which are extremely small today, but do still exist, are the uncertainties concerning the definitions of the units.

10.4 Energy Density in the Electromagnetic Field

In Sect. 3.5, we discussed a number of different expressions of the energy of an electrostatic system. All of them are equivalent in electrostatics. Similarly, in Sect. 8.4, we discussed a number of expressions of the energy of a system of steady currents, equivalent under time-independent conditions. Under dynamical conditions, only the expressions of energy as the energy of the fields remain valid, namely

$$U_E = \frac{\epsilon_0}{2} \int_{\text{all space}} \mathbf{E}^2 dV, \quad U_M = \frac{1}{2\mu_0} \int_{\text{all space}} \mathbf{B}^2 dV,$$

as we shall now prove. We shall prove that the energy per unit volume in an electromagnetic field under any set of conditions is

$$w = \frac{\epsilon_0}{2} \mathbf{E}^2 + \frac{1}{2\mu_0} \mathbf{B}^2.$$

Let us start with an electromagnetic field in vacuum, namely in the absence of matter. We assume local energy conservation to hold, namely that energy is not only conserved, but that it is conserved locally. This means that if the energy contained in a generic (small) volume ΔV diminishes in a second by a certain quantity, namely

$$-\frac{d}{dt} \int_{\Delta V} w dV = - \int_{\Delta V} \frac{\partial w}{\partial t} dV,$$

the same quantity of energy must exit in a second through the surface Σ that bounds ΔV . Let \mathbf{S} be the vector having a magnitude equal to the energy flux through the unit's surface normal to the flow direction per unit time and the direction and sense of the flux. The energy flowing through Σ in a second is then

$$\int_{\Sigma} \mathbf{S} \cdot \mathbf{n} d\Sigma \quad (10.31)$$

and we can express in a formula what we have just stated in words as

$$-\int_{\Delta V} \frac{\partial w}{\partial t} dV = \int_{\Sigma} \mathbf{S} \cdot \mathbf{n} d\Sigma, \quad (10.32)$$

where \mathbf{n} is the outgoing normal unit vector. We now use the divergence theorem to transform the surface integral on the right-hand side into a volume integral on ΔV . Being that the volume is arbitrary, the equality must hold for the integral, namely

$$\nabla \cdot \mathbf{S} = -\frac{\partial w}{\partial t}. \quad (10.33)$$

We already encountered this equation expressing the local conservation of the electric charge. We now have the continuity equation for energy. Equation (10.33) holds in a vacuum, when the electromagnetic field does not exchange energy with matter. In the presence of matter, namely if charged particles are present, we must take this energy exchange into account.

We now state that the energy decrease in ΔV in a second is equal to the energy flowing out in a second plus the work done by the field on the charges in that second.

Let \mathbf{j} be the current density, q the charge of the particles (assumed to be uniformly equal, for simplicity), n_p the number of charges per unit volume and \mathbf{v}_i the velocity of the i -th charge. Equation (5.20) gives the work done as being the electric field per unit time in the volume dV . Note that we demonstrated this equation when discussing a time-independent electric field. However, we did not use this hypothesis in the demonstration, which is consequently valid under time-dependent conditions as well. We now also have a magnetic field, but its work on the charges is zero, because the Lorentz force is always perpendicular to the velocity of the charge. We can then say that the work done by the electromagnetic field in a second on the charges in dV is $\mathbf{E} \cdot \mathbf{j} dV$.

The energy balance is given by

$$-\int_{\Delta V} \frac{\partial w}{\partial t} dV = \int_{\Sigma} \mathbf{S} \cdot \mathbf{n} d\Sigma + \int_{\Delta V} \mathbf{E} \cdot \mathbf{j} dV. \quad (10.34)$$

As we did above, we use the divergence theorem to transform the surface integral into a volume integral and extend the resulting equality to the integrands, given that the integration volume is arbitrary. The resulting continuity equation is

$$\mathbf{E} \cdot \mathbf{j} = -\frac{\partial w}{\partial t} - \nabla \cdot \mathbf{S}. \quad (10.35)$$

We now need to find the expressions of the electromagnetic energy density w and energy flux \mathbf{S} . To that aim, we search for two functions of the fields, only one scalar and one vector that satisfy Eq. (10.35). Let us start from the fourth Maxwell equation

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}.$$

We solve it for \mathbf{j} and take the scalar product with \mathbf{E} of both sides, obtaining

$$\mathbf{j} \cdot \mathbf{E} = \frac{1}{\mu_0} \mathbf{E} \cdot \nabla \times \mathbf{B} - \epsilon_0 \mathbf{E} \cdot \frac{\partial \mathbf{E}}{\partial t}.$$

We now use the vector identity

$$\mathbf{E} \cdot \nabla \times \mathbf{B} = \mathbf{B} \cdot (\nabla \times \mathbf{E}) + \nabla \cdot (\mathbf{B} \times \mathbf{E})$$

obtaining

$$\mathbf{j} \cdot \mathbf{E} = \frac{1}{\mu_0} [\mathbf{B} \cdot (\nabla \times \mathbf{E}) + \nabla \cdot (\mathbf{B} \times \mathbf{E})] - \frac{\partial(\epsilon_0 \mathbf{E}^2 / 2)}{\partial t}.$$

Using the Faraday law $\nabla \times \mathbf{E} = -\partial \mathbf{B} / \partial t$, and inverting the factors in the cross product of the fields, we write

$$\mathbf{j} \cdot \mathbf{E} = -\nabla \cdot \left(\frac{1}{\mu_0} \mathbf{E} \times \mathbf{B} \right) - \frac{\partial}{\partial t} \left(\frac{\epsilon_0}{2} \mathbf{E}^2 + \frac{1}{2\mu_0} \mathbf{B}^2 \right),$$

which has the form we have been searching for, with energy density

$$w = \frac{\epsilon_0}{2} \mathbf{E}^2 + \frac{1}{2\mu_0} \mathbf{B}^2 \quad (10.36)$$

and energy flux

$$\mathbf{S} = \frac{1}{\mu_0} \mathbf{E} \times \mathbf{B}. \quad (10.37)$$

The just-demonstrated theorem is credited to John Henry Poynting (UK, 1852–1914), who published his result in 1864, and is called the Poynting theorem. The vector \mathbf{S} is called the Poynting vector. We here notice that the Poynting theorem shows that the above equations are *possible* expressions of the electromagnetic field energy density and energy flux, but it does not demonstrate them to be the unique solutions. As a matter of fact, other solutions do exist, namely other functions of the electric and magnetic fields alone, one scalar and one vector satisfying Eq. (10.34). All of them, however, are more complex than Eqs. (10.36) and (10.37) and contain derivatives of the fields. The choice universally made of Eqs. (10.36) and (10.37) is uniquely based on a simplicity criterion.

We now give the expressions of the energy density and energy flux in the presence of material media, both dielectric and magnetic. We shall assume the materials to be *linear*, namely \mathbf{D} to be proportional to \mathbf{E} and \mathbf{H} to be proportional to \mathbf{B} . We found the energy density of the fields in electrostatics in Eq. (4.53), which we now generalize to dynamic conditions. The demonstration is quite similar to that we gave for a vacuum and we shall not give it here, for the sake of brevity. The results are the equations

$$w = \frac{1}{2} \mathbf{E} \cdot \mathbf{D} + \frac{1}{2} \mathbf{B} \cdot \mathbf{H} \quad (10.38)$$

and

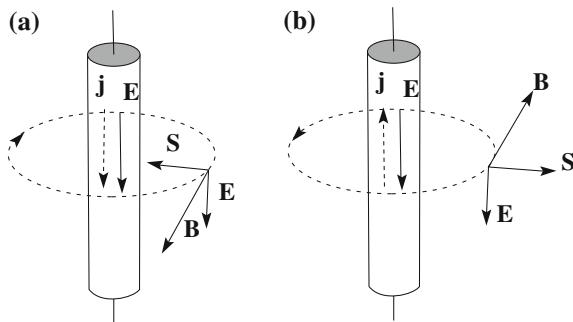
$$\mathbf{S} = \mathbf{E} \times \mathbf{H}. \quad (10.39)$$

10.5 Energy Flux

In this section, we shall consider a few examples of energy flux to familiarize ourselves with the concept. These examples show that a non-zero energy flux, in addition to the energy density, also exists when the fields do not depend on time. Under dynamic conditions, the Poynting vector is even more important. We shall treat that in the 4th volume of the course, after having studied the relevant properties of wave phenomena.

In the first example, we consider a simple cylindrical resistor R carrying a continuous current I , as shown in Fig. 10.9a. A battery is used to apply the potential difference $V = RI$ between the ends of the resistor. For simplicity, we assume the battery to be cylindrical, as is the resistor.

Fig. 10.9 Current density, electric and magnetic field and Poynting vector in **a** a resistor, **b** a battery



An electric field \mathbf{E} is present inside the resistor and in the immediate neighboring vicinity, which is directed substantially parallel to the resistor. The current produces a magnetic field \mathbf{B} , whose lines are circles normal to the resistor and centered on its axis. Hence, the Poynting vector \mathbf{S} is radially directed towards the axis. As we know, the resistor dissipates energy through the Joule effect. This energy is provided by the battery and one might think that it moves from the battery to the resistor together with the current, flowing through the connecting wires. But this is not so. As we have seen, energy enters the resistor from the space outside through the lateral surface.

Let us now consider the battery, shown in Fig. 10.9b. In the battery, the current runs in the opposite direction relative to the electric field \mathbf{E} . Compared to the previously-analyzed situation, the directions of \mathbf{E} and \mathbf{B} are the same and opposite, respectively. Consequently, \mathbf{S} is now directed normally outside the battery axis. The energy produced by the battery flows out through its lateral surface, advances in the empty space, reaches the resistor and enters it through its lateral surface. The energy in any space volume between battery and resistor is constant in time, because the fields are constant. But the volume is continuously crossed by energy, which flows in from one side and out the other in the same amount.

In order to understand the process better, let us think of the effect of the battery as acting on pairs of a positive and negative charge, taking them apart and pushing them in opposite directions through the wires connecting the ends of the battery with the ends of the resistor. The pair of charges finally comes back together in the resistor. Initially, the two charges are close to one another inside the battery. Their field is mainly between them, in a small region of space. Consequently, the field energy is localized inside the battery as well. When the charges separate, the lines of their field open up, invading greater and greater regions of space outside the battery (moving mainly through the lateral surface). Contemporarily, the field energy becomes distributed through increasingly large regions of empty space, clearly with diminishing energy density. Then, when the two charges once again approach one another inside the resistor, their fields condense as well and so does the energy, which finally degrades into thermal energy.

The Maxwell theory leads to unexpected conclusions concerning the energy flux. Conductors appear to be such for an electric current, but not for energy. On the

contrary, conductors destroy macroscopic energy, transforming it into thermal energy. Vacuum and dielectrics, which are charge insulators, are energy conductors.

Let us now consider a second example, which is a quasi-static one. Let us consider a parallel plate capacitor, as shown in Fig. 10.10, while it is being charged by a current produced by a battery. The process is time-dependent, but we assume the current intensity to vary very slowly, namely slow enough to be practically equal in all the sections of the circuit at any given instant. In the process, energy is transferred from the battery to the capacitor, namely to its electric field.

Let r_0 be the radius of the circular plates and h their distance. Neglecting the fringe effects, the energy of the capacitor at the generic instant is $U = \frac{\epsilon_0}{2} E^2 \pi r_0^2 h$. Its rate of change is

$$\frac{dU}{dt} = \epsilon_0 E \frac{dE}{dt} \pi r_0^2 h.$$

The time variation of the electric field gives origin to a magnetic field \mathbf{B} , whose lines are circles centered on the axis, as shown in Fig. 10.10. We apply the Ampère law, generalized to include the displacement current, to the circle at the rim of the plates, as in the figure, obtaining.

$$2\pi r_0 B = \mu_0 \epsilon_0 \frac{dE}{dt} \pi r_0^2,$$

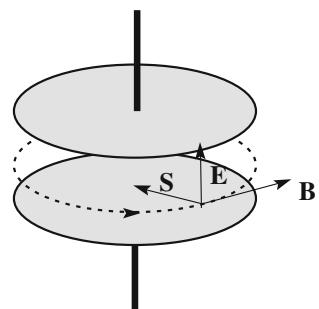
from which we have the magnitude of the field

$$B = \frac{\mu_0 \epsilon_0}{2} r_0 \frac{dE}{dt}$$

The Poynting vector is directed radially. Its sense, in the case we are considering in which E is growing, is toward the inside. In this case too, energy enters the capacitor through its lateral surface.

Let us check that the total flux entering in a second is equal to the rate of change of the capacitor's energy. The energy flux is the Poynting vector magnitude times the surface area, which is $2\pi r_0 h$, and we write

Fig. 10.10 Electric and magnetic fields and Poynting vector in a parallel plate capacitor during charging



$$S2\pi r_0 h = \frac{EB}{\mu_0} 2\pi r_0 h = \varepsilon_0 E \frac{dE}{dt} \pi r_0^2 h,$$

which is equal to the above-found expression of dU/dt .

QUESTION Q 10.4. Consider a cylindrical solenoid having length l , radius r and n turns per unit length carrying a current varying with time of intensity $I(t)$. Neglect the fringe effects. Calculate the magnetic field, the electric field and the Poynting vector, in magnitude and direction, immediately inside the solenoid. Calculate the energy stored in the solenoid and its rate of change and compare with the energy flux through the lateral surface. \square

10.6 Momentum Density in the Electromagnetic Field

In the study of physics, one quite often encounters situations in which the energy of an isolated system appears not to be conserved. However, in every case, the energy non-conservation is only apparent. A closer analysis always shows that energy has not been lost, but rather has transformed into another type, which we had not taken into account. In mechanics, for example, we observe that the kinetic energy of a body moving along a horizontal plane decreases with time. However, if we measure its temperature, we find it increased. The (macroscopic) kinetic energy has transformed into (microscopic) thermal energy. The sole mechanical energy is not conserved, but the total energy is constant. The mechanical energy of a system of charged bodies is not constant, because they exchange energy with the electromagnetic field. What is constant is the sum of the field and the mechanical energies. In conclusion, energy conservation is a universal law without exceptions. When energy appears not to be conserved, it is because it has changed into a “hidden” form that we forgot to account for.

The conservation laws of the linear and angular momenta of an isolated system are completely general as well. As for the energy, however, it happens that linear momentum and angular momentum may change to other forms, of which we might not have known in advance. Having neglected to include these forms in the balance, those quantities appear not to be conserved. In the study of mechanics and thermodynamics, we encounter cases of apparent non-conservation of energy, but never cases of apparent non-conservation of linear and angular momenta. As opposed to energy, the latter are vector quantities, a characteristic that makes them more difficult to hide. They “succeed” in doing so in systems of electrically-charged bodies in motion. In other words, neither the total mechanical linear momentum nor the total mechanical angular momentum of an isolated system of charges are conserved.

Consider, for example, the simple system consisting of two point-like charged particles in motion. Each of them acts on the other with a force (which, we notice, is not the Coulomb force if the charges are moving). Obviously, the two forces are an action and reaction pair and are the only forces acting on the system. We would

consequently expect the two forces to be equal and opposite (resultant force equals zero) acting on the same line (resulting moment equals zero) or, in an equivalent manner, that the total linear and angular momenta are conserved. Well, neither the first expectation nor the second are confirmed by experiments.

These non-conservations are, however, only apparent and are due to our not having included in the balance the field linear momentum in one case and the field angular momentum in the other. Indeed, the electromagnetic field not only contains energy, but linear and angular momenta as well. These quantities, as is the case for energy, are distributed throughout the entire space. We talk of linear momentum density and angular momentum density, which are the quantities per unit volume. In the evolution of a system of charged bodies, linear and angular momenta are continuously exchanged between the bodies and the field.

Let us start with an example of linear momentum. Let us recall that any volume dV in the field stores an energy wdV , where w is the function of the fields given by Eq. (10.36). As we have already discussed, matter and field exchange energy. The total energy, the sum of the two forms, is not only conserved, but is locally conserved. If we want linear momentum to be conserved as well, we must admit that the field also has linear momentum. More specifically, any infinitesimal volume dV should have a linear momentum proportional to dV , say $\mathbf{g}(x, y, z, t)dV$, where $\mathbf{g}(x, y, z, t)$ is a vector quantity, a function of the coordinates and time, which is the *linear momentum density*. Matter and field can exchange linear momentum and we shall assume that their sum is locally conserved. We shall also assume, as we did for energy, that \mathbf{g} can be expressed in terms of the fields alone. To find this expression, one proceeds in a manner similar to what we did for the energy density and the energy flux in Sect. 10.3. An additional complication comes from the fact that \mathbf{g} is a vector, rather than a scalar like w . One separately considers each of its components, for example g_x . One then imposes that the variation in the time interval dt of g_x in an arbitrary infinitesimal dV , namely $g_x dV$, should be equal to the x -component of linear momentum that flows out in dt through the surface surrounding dV , decreased by the x -component of linear momentum transferred from the field to matter. To write this equation, we must introduce a physical quantity expressing the flux of the x -component of the field linear momentum. This quantity has three components. Clearly, there are two other three-components quantities for the fluxes of the y and z components of the field momentum. This nine-component object is mathematically a tensor (you can think of it as a 3×3 matrix). This complication apart, the argument proceeds exactly as it did for energy and its flux. We look for two expressions in terms of the fields, a vector (the linear momentum density) and a tensor (the linear momentum flux) satisfying the continuity equation (namely the local momentum conservation). We shall not go through the details but will give only the result for the momentum density. However, we observe that the very fact that a result is obtained shows that the Maxwell theory foresees linear momentum conservation. Only experiments can prove, however, whether this is true. And indeed, experiments have shown that linear momentum is conserved.

The result is that the *field* momentum density is simply the Poynting vector divided by the square of the light velocity in vacuum, namely

$$\mathbf{g} = \mathbf{S}/c^2 = \varepsilon_0\mu_0\mathbf{S} = \varepsilon_0\mathbf{E} \times \mathbf{B}. \quad (10.40)$$

Let us consider the simple example of two point charges q_1 and q_2 in motion with velocities \mathbf{v}_1 and \mathbf{v}_2 , respectively. The charges being in motion, the forces \mathbf{F}_{12} exerted by q_1 on q_2 and \mathbf{F}_{21} exerted by q_2 on q_1 are not the Coulomb force. We do not know the expression of the force, which is quite complicated for a general motion. The two forces are not necessarily equal and opposite. As we have just seen, the linear momentum of the charges is not always conserved. Consequently, the action and reaction law may not hold for electromagnetic forces.

If we remember that no information can propagate in space at a speed greater than the velocity of light, we understand that the action and reaction law not only may, but must be violated. Indeed, in an interval dt at a certain instant of time, the momentum of, say, q_1 varies under the action of the force \mathbf{F}_{21} . The action and reaction law would require that the momentum of q_2 should vary under the action of \mathbf{F}_{21} *contemporarily* in dt , by an equal and opposite quantity in order to guarantee the conservation of linear momentum in every instant. But this cannot be true, because the information that v_1 has changed must reach q_2 moving at the speed of light. This velocity is very large, but finite, and the propagation requires a non-zero time. During this time interval, the momentum change is, so to speak, traveling between the charges, and the total linear momentum of the two charges has varied. Clearly, if the two charges were at rest and had been so for a long enough time, the field would have been constant during this time and the information on the relative position of the charges would have had time to propagate from one to the other. The Coulomb law and the action and reaction law hold.

In conclusion, the action and reaction law is not a universal law of physics. It does not hold in every case. The fundamental laws are the linear and angular momenta conservations. The action and reaction law is a consequence of these two laws and is valid only when the momentum and angular momentum of the field are constant.

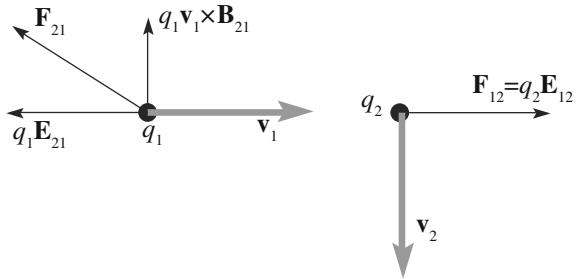
We shall now discuss an example of the non-conservation of matter linear momentum and one of the non-conservation of matter angular momentum.

Violation of the action and reaction law.

Let us consider two point charges q_1 and q_2 in motion with velocities \mathbf{v}_1 and \mathbf{v}_2 , respectively, perpendicular to one another. Let us consider the instant in which the second particle crosses the road in front of the first one, as shown in Fig. 10.11. The system is isolated, namely no external forces act on the particles. The internal forces between the charges are due to the electric and magnetic fields each of them produces in the position of the other. We have not studied the electric and magnetic fields produced by charges in motion. To find them, one starts from the fields of a point charge at rest, which we know, and applies the Lorentz transformations to them.

We shall not deal with that but only give the result. The electric field of a point charge in rectilinear uniform motion is directed away from or toward the charge, if it is positive or negative, respectively, as would a charge at rest. However, as opposed to a charge at rest, the field intensity depends not only on the distance

Fig. 10.11 Fields and mutual forces of two point charges in motion



r (still $1/r^2$), but also on the angle between the position vector \mathbf{r} from the charge to the considered point with the velocity of the charge. At the same r , the field intensity is largest in the plane perpendicular to velocity, the smallest in the direction of velocity (both forward and backward). For the arguments we are considering here, it is sufficient to know that the field is central.

A charge in rectilinear uniform motion produces a magnetic field too. Indeed, a straight continuous electric current is only made of charges moving in such a motion. We can then use what we know to find the field of a charge of velocity \mathbf{v} . Recalling that the speeds of the charge carriers are much smaller than the speed of light, we shall find an expression valid under these conditions. Equation (6.52) gives the magnetic field produced by a current density \mathbf{j} in the generic point P_2 . We reproduce it here for convenience:

$$\mathbf{B}(P_1) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{j}(P_2) \times \mathbf{r}_{21}}{r_{21}^3} dV_2,$$

where \mathbf{r}_{21} is the vector from the integration point P_2 to the point at which we evaluate the field, P_1 . The expression holds for constant current intensity, corresponding to constant velocity of the carriers. We can interpret this expression by saying that the generic volume ΔV contributes to the field with

$$\Delta \mathbf{B}(P_1) = \frac{\mu_0 \mathbf{j}(P_2) \times \mathbf{r}_{21}}{4\pi r_{21}^3} dV_2.$$

On the other hand, $\mathbf{j}\Delta V$ is the sum of the contributions of the single carriers in ΔV . If n_p is the number of carriers per unit volume and \mathbf{v}_i are their velocities, we have

$$\mathbf{j}\Delta V = qn_p \Delta V \langle \mathbf{v} \rangle = q \sum_{\Delta V} \mathbf{v}_i.$$

The contribution $\Delta\mathbf{B}$ of the volume ΔV is then

$$\Delta\mathbf{B}(P_1) = \sum_{\Delta V} \mathbf{B}_i(P_i) = \frac{\mu_0 q}{4\pi} \sum_{\Delta V} \frac{\mathbf{v}_i \times \mathbf{r}_{21}}{r_{21}^3}$$

We recognize in this expression that the magnetic field produced in P_1 by the point charge q passing in P_2 with velocity \mathbf{v} is

$$\mathbf{B}(P_1) = \frac{\mu_0 q}{4\pi} \frac{\mathbf{v} \times \mathbf{r}_{21}}{r_{21}^3}. \quad (10.41)$$

This is the expression we need. We repeat that it is valid for velocity much smaller than the speed of light.

The lines of the magnetic field of Eq. (10.41) are, as expected, circles normal to the velocity centered on the trajectory of the charge. The field intensity is larger for higher velocities.

We can now analyze the forces exchanged by the two charges in our example. The force \mathbf{F}_{12} exerted by q_1 on q_2 is purely electric, because the magnetic field of q_1 is zero along the straight line of its trajectory (both in front and behind q_1), because \mathbf{v} and \mathbf{r}_{21} are parallel at these points. Consequently, \mathbf{F}_{12} is directed as the line joining the charges, as shown in Fig. 10.11.

Contrastingly, the force \mathbf{F}_{21} exerted by q_2 on q_1 has electric and magnetic components. The electric component is $q_1 \mathbf{E}_{21}$, where \mathbf{E}_{21} is the electric field of q_2 in the position of q_1 . It has the direction of the line joining the charges. The magnetic field \mathbf{B}_{21} of q_2 in the position of q_1 has the direction normal to the drawing to inside the page. The corresponding force on q_1 is perpendicular to its velocity to the left.

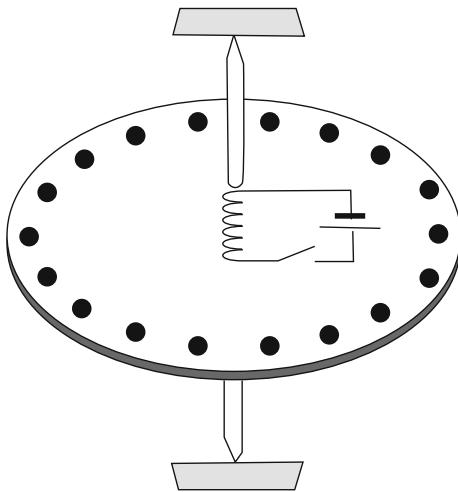
We see that \mathbf{F}_{12} and \mathbf{F}_{21} are *not* equal and opposite with the same application line. In other words, the total mechanic linear momentum of the system of two charges is not constant in time, even if no external force is acting on the system. However, the field linear momentum is also changing in such a way, as we might calculate, that the total (matter plus field) linear momentum remains constant. The linear momentum conservation holds only if we take into account both contributions.

The Feynman paradox.

Richard Feynman (US, 1918–1988) gave a beautiful example of electromagnetic field angular momentum in his Lectures on Physics in 1963. We shall discuss it in detail now.

The device in Fig. 10.12 is a plastic disk free to turn about its vertical axis through the center. N metal spheres of electric charges Q are equally spaced and rigidly fixed about its rim on a circle Γ of radius R . A solenoid on the symmetry axis initially carries the steady current I . At the beginning, the system is at rest. The current produces a magnetic field \mathbf{B} . The flux of \mathbf{B} is largely linked to Γ . At a given instant, we open the circuit (we imagine doing that with an internal action). The current and the magnetic field decrease to zero.

Fig. 10.12 A rigid disk free to turn on its vertical central axis. The black dots on its rim are equal electric charges



If Σ is the circle inside Γ , the initial flux of \mathbf{B} through the circle Σ is

$$\Phi_B = \int_{\Sigma} \mathbf{B} \cdot \mathbf{n} d\Sigma. \quad (10.42)$$

During the shutdown, the flux of \mathbf{B} decreases with time and an induced emf appears about Γ . This is similar to the betatron case we analyzed in Sect. 7.4. Repeating the argument, we can say that the electric field \mathbf{E} induced at the points of Γ is such that

$$\frac{d\Phi_B}{dt} = - \oint_{\Gamma} \mathbf{E} \cdot d\mathbf{s} = -2\pi R E. \quad (10.43)$$

The (variable with time) force QE is now acting on each sphere. Its direction is tangent to Γ . Consequently, the disk starts a rotation about the axis, with increasing angular velocity, which reaches its final value when the current, and the magnetic field, have vanished. The angular momentum of the mechanical system, which was initially zero, is now something other than zero. This happened under the action of internal forces alone. The moment of external forces has been zero. It looks like the angular momentum conservation law has been violated. This, however, is true only if we consider the mechanic angular momentum by itself. It is the sum of the angular momentum of the disk and the electromagnetic field that remains constant, as we shall now see.

Let us go back to the initial state of the system. The fields at large distances from the disk are easier to calculate. At such distances, the magnetic field \mathbf{B} is the field of a magnetic dipole, and the electric field \mathbf{E} can be approximated with a monopole, considering all the charges to be a point charge NQ in the center of the disk.

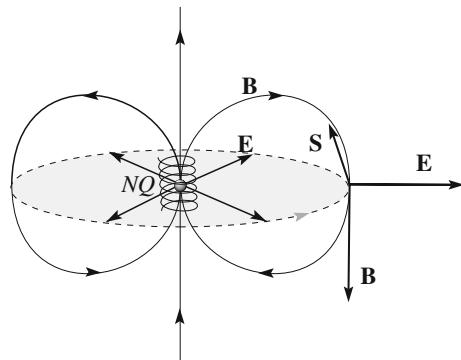
The lines of the Poynting vector \mathbf{S} are consequently circumference centered on the axis. One of these is shown in Fig. 10.13. The lines of \mathbf{S} are the lines along which the energy of the field flows in the initial state of the system. When everything is steady, every infinitesimal space volume contains a constant amount of energy and is continually crossed by a constant flux of energy. This situation, which looks absurd at first sight, is far from being so if carefully analyzed. Indeed, Eq. (10.40) tells us that the energy flux is proportional to the linear momentum density of the field. And a circular flow of linear momentum corresponds to an angular momentum about the axis. The experiment we have described shows that the angular momentum can be conserved only if the angular momentum acquired by the mechanical system was initially present in the field. The energy flux in the stationary field, rather than being an absurdity, is absolutely necessary.

Summarizing, we can say that the described experiment shows that the disk, which initially has no angular momentum, acquires angular momentum when the current is shut down. No force external to the disk plus the field system is present. The angular momentum conservation is satisfied by the angular momentum initially stored in the field, in the closed lines of the linear momentum density \mathbf{g} .

Let us now quantitatively analyze the problem. We start by calculating the magnetic flux Φ_B in Eq. (10.42) linked to the circumference Γ (radius R) through the spheres. Let us consider the plane of the spheres. Given that the lines of \mathbf{B} are closed, the flux of \mathbf{B} through Γ is equal and opposite to the flux of \mathbf{B} through the part of the plane external to Γ (which we call Σ_{ext}) can be more easily calculated because its points are at distances large compared to the dimensions of the solenoid. We can say that \mathbf{B} is the field of a magnetic dipole of magnetic moment μ (proportional to the current intensity). Its expression in the equatorial plane is particularly simple, given by the third of Eq. (6.61) for $z = 0$. At the distance r from the axis, \mathbf{B} is

$$\mathbf{B} = -\frac{\mu_0 \mu}{4\pi r^3} \cdot \hat{\mathbf{r}} \quad (10.44)$$

Fig. 10.13 The fields during the decrease of the solenoid current



We now calculate its flux, taking as surface elements the annuluses of radii r and $r + dr$, obtaining

$$\Phi_B = - \int_{\Sigma_{\text{ext}}} \mathbf{B} \cdot \mathbf{n} d\Sigma = \frac{\mu_0}{4\pi} \int_R^\infty \frac{\mu}{r^3} 2\pi r dr.$$

We then obtain by integration

$$\Phi_B = \frac{\mu_0 \mu}{2 R}. \quad (10.45)$$

The magnitude of the induced electric field \mathbf{E} acting on the spheres is then (by the flux rule)

$$E = -\frac{1}{2\pi R} \frac{d\Phi_B}{dt} = -\frac{\mu_0}{4\pi} \frac{1}{R^2} \frac{d\mu}{dt}.$$

The torque on the disk is then

$$M = ENQR = -\frac{\mu_0}{4\pi} \frac{NQ}{R} \frac{d\mu}{dt}.$$

Let \mathbf{L} be the mechanical angular momentum. Its rate of change is $\mathbf{M} = d\mathbf{L}/dt$, and we obtain the final value of L after the current has vanished by calculating its variation from the initial value, which is zero, to the final value, by integration, namely

$$L = \int_0^\infty \frac{dL}{dt} dt = -\frac{\mu_0}{4\pi} \frac{NQ}{R} \int_0^\infty \frac{d\mu}{dt} dt.$$

Now, considering that the magnetic moment varies from the initial value μ to zero, we obtain

$$L = \frac{\mu_0}{4\pi} \frac{NQ}{R} \mu. \quad (10.46)$$

Let us now analyze the angular momentum stored in the initial electromagnetic field. The easiest way to calculate this quantity is to think of assembling the system starting from a state in which the current is present but the charges are at infinite distances and do not interact with it. Let us bring a small fraction δQ of each of the charges from infinity. We proceed by respecting the symmetry of the problem, moving each charge element radially into its final position along straight lines in the equatorial plane. The force on each δQ is the Lorentz force

$$\delta\mathbf{F} = \delta Q(\mathbf{E} + \mathbf{v} \times \mathbf{B}), \quad (10.47)$$

where the charge velocity \mathbf{v} is directed radially toward the axis. The field \mathbf{E} is directed radially outward and has an intensity dependent on how much charge is already in place. The magnetic field \mathbf{B} is due to the solenoid and, on the equatorial plane, is given by Eq. (10.44). The torque about the center on δQ is

$$\delta\mathbf{M} = \delta Q\mathbf{r} \times (\mathbf{v} \times \mathbf{B}). \quad (10.48)$$

Note that the electric part of the Lorentz force does not contribute to the moment, because the displacements, hence the velocities, of the charges δQ are parallel to \mathbf{r} . To move each δQ , we must apply a torque equal and opposite to $\delta\mathbf{M}$. The application of $-\delta\mathbf{M}$ in a time interval dt produces a change of the (mechanical) angular momentum of the N spheres system of $d(\delta L) = -N\delta M dt = -N\delta Q v B(r) dt$, where we have taken into account that the vectors \mathbf{r} , \mathbf{v} and \mathbf{B} are mutually perpendicular. $B(r)$ is the magnitude of the field given by Eq. (10.44). Notice now that $v dt$ is equal to the displacement dr of the charge δQ in dt . Integrating on the complete displacements of the N charges, we get the angular momentum variation corresponding to the translation of δQ from infinity, namely

$$\delta L = \int d(\delta L) = -N\delta Q \int_{\infty}^R r B dr = -N\delta Q \frac{\mu_0}{4\pi} \mu \int_{\infty}^R r^{-2} dr = \frac{\mu_0 \mu}{4\pi R} N\delta Q.$$

Adding up the contributions of all the charge elements δQ , we finally obtain

$$L = \frac{\mu_0 N Q}{4\pi R} \mu,$$

which is equal to Eq. (10.46).

10.7 Maxwell Equations in Matter

The theoretical and experimental work of James Clerk Maxwell that we have described in this chapter resulted in the unification of electricity, magnetism and optics, which are all ruled by the Maxwell equations. We write the four equations again, both in the differential and integral forms:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}, \quad \int_{\Sigma} \mathbf{E} \cdot \mathbf{n} d\Sigma = \frac{Q_{\text{int}}}{\epsilon_0}, \quad (10.49)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad \oint_{\Gamma} \mathbf{E} \cdot d\mathbf{s} = -\frac{d\Phi_B}{dt}, \quad (10.50)$$

$$\nabla \cdot \mathbf{B} = 0, \quad \int_{\Sigma} \mathbf{B} \cdot \mathbf{n} d\Sigma = 0, \quad (10.51)$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}, \quad \oint_{\Gamma} \mathbf{E} \cdot d\mathbf{s} = \mu_0 I_{\text{link}} + \mu_0 \varepsilon_0 \frac{d\Phi_E}{dt}. \quad (10.52)$$

Maxwell equations include the local electric charge conservation.

The Maxwell equations are read as follows.

The first equation states that the sources and the sinks of the electric field are the electric charges or, in integral terms, that the flux of the electric field flowing out from any closed surface is equal to the total charge inside the surface divided by ε_0 .

The second equation states that a magnetic field variable with time gives origin to an electric field. In integral terms, the electric field circulation about any closed line is equal to the opposite of the rate of change of the magnetic field flux linked to that line.

The third equation states that sources and sinks of the magnetic field do not exist, or, in integral terms, that the flux of the magnetic field flowing out from any closed surface is zero.

The fourth equation states that the magnetic field has two origins: charge currents and an electric field variable with time. In integral terms, the magnetic field circulation about any closed line is equal to μ_0 times the linked to the line current intensity plus $\varepsilon_0 \mu_0$ times the rate of change of the electric field linked flux.

We recall that the charge and current densities in Eqs. (10.49) and (10.50) include, respectively, all the charges and currents, both of macroscopic and microscopic nature. As we learned, however, we directly control only the macroscopic quantities. In the presence of materials, it is, consequently, convenient to rewrite the Maxwell equations in terms of the free charges density ρ_f and the conduction current density \mathbf{j}_c . We did that under time-independent conditions in Chaps. 4 and 9, respectively. Let us proceed now under general conditions.

The first part of the argument for the first equation is exactly the same as that for electrostatics. We summarize it for convenience. We start by expressing the charge density as the sum of two terms, one the result of the free charges, the other the result of the polarization charges, namely $\rho = \rho_f + \rho_p$. We then observe that the polarization charge is linked to the polarization density by the equation $\rho_p = -\nabla \cdot \mathbf{P}$, and we introduce the electric displacement auxiliary field

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P} \quad (10.53)$$

and re-write Eq. (10.46) as

$$\nabla \cdot \mathbf{D} = \rho_f, \quad \int_{\Sigma} \mathbf{D} \cdot \mathbf{n} d\Sigma = Q_{f,\text{int}}, \quad (10.54)$$

As for the fourth equation, in magnetostatics, we considered two contributions to the current density \mathbf{j} , namely the macroscopic conduction current density \mathbf{j}_c and the microscopic magnetization current density \mathbf{j}_m . The latter, also under dynamic conditions, is linked to the magnetization density \mathbf{M} by the equation

$$\mathbf{j}_m = \nabla \times \mathbf{M} \quad (10.55)$$

Under dynamic conditions, we must take into account that the polarization charges move in time, producing a second microscopic current. To find its expression, consider a unitary surface inside the dielectric medium. Considering that the polarization vector \mathbf{P} is the dipole moment per unit volume and that a dipole moment is a charge times a distance, it is easy to understand that the charge crossing the unitary surface in a second is equal to the rate of change of \mathbf{P} . The polarization current density is then

$$\mathbf{j}_p = \frac{d\mathbf{P}}{dt} \quad (10.56)$$

We then re-write Eq. (10.52) as

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j}_c + \mu_0 \mathbf{j}_m + \mu_0 \mathbf{j}_p + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}$$

and, using Eqs. (10.55) and (10.56),

$$\nabla \times \left(\frac{\mathbf{B}}{\mu_0} - \mathbf{M} \right) = \mathbf{j}_c + \frac{\partial(\mathbf{P} + \epsilon_0 \mathbf{E})}{\partial t},$$

where we recognize the known auxiliary fields

$$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{M} \quad (10.57)$$

on the left-hand side and \mathbf{D} on the right-hand side. We can then write the fourth Maxwell equation as

$$\nabla \times \mathbf{H} = \mathbf{j}_c + \frac{\partial \mathbf{D}}{\partial t}, \quad \oint_{\Gamma} \mathbf{H} \cdot d\mathbf{s} = I_{\text{link}} + \frac{d\Phi_D}{dt}. \quad (10.58)$$

We do not need to change anything in the second and third equations, because they do not contain charges or currents. The Maxwell equations can be summarized in terms of macroscopic charges and currents with the following six equations, which we repeat for convenience, in the fields \mathbf{E} , \mathbf{D} , \mathbf{B} , \mathbf{H} , \mathbf{P} and \mathbf{M}

$$\nabla \cdot \mathbf{D} = \rho_f, \quad \int_{\Sigma} \mathbf{D} \cdot \mathbf{n} d\Sigma = Q_{f,\text{int}}, \quad (10.59)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad \oint_{\Gamma} \mathbf{E} \cdot d\mathbf{s} = -\frac{d\Phi_B}{dt}, \quad (10.60)$$

$$\nabla \cdot \mathbf{B} = 0, \quad \int_{\Sigma} \mathbf{B} \cdot \mathbf{n} d\Sigma = 0, \quad (10.61)$$

$$\nabla \times \mathbf{H} = \mathbf{j}_c + \frac{\partial \mathbf{D}}{\partial t}, \quad \oint_{\Gamma} \mathbf{H} \cdot d\mathbf{s} = I_{\text{link}} + \frac{d\Phi_D}{dt}. \quad (10.62)$$

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \quad (10.63)$$

$$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{M} \quad (10.64)$$

We recall here once more that the quantities in the above equations, namely charge, charge density, current, current density and the fields \mathbf{E} and \mathbf{B} , are macroscopic quantities, namely they are average values taken on volumes much smaller than the macroscopic dimensions but still much larger than the molecular ones (namely $\gg 0.1$ nm) and on time intervals much smaller than those characteristic of macroscopic phenomena but still much larger than the molecular and atomic ones (namely \gg femtoseconds). The fields \mathbf{D} and \mathbf{H} are purely macroscopic auxiliary fields, with no counterpart at the level of the constituents of matter.

On the other hand, it is evident now, as it was under time-independent conditions, that the above equations are insufficient. In addition to them, we need the relations linking the polarization \mathbf{P} to the electric field \mathbf{E} and the magnetization \mathbf{M} to the magnetic field \mathbf{B} . The problem is intrinsically complex, because the searched-for relations should describe the reaction of matter to the action of electric and magnetic fields. It was complex under static conditions; it is even more complex under time-dependent ones.

Let us recall that, under static conditions, and within the limits we shall recall, we have

$$\mathbf{P} = \epsilon_0 \chi_e \mathbf{E}, \quad \mathbf{M} = \frac{\chi_m'}{\mu_0} \mathbf{B}.$$

These proportionality relations hold for linear media, namely when the response of matter to the electric or magnetic solicitation is proportional to the solicitation. The second expression, in particular, does not hold for ferromagnetic media. In addition, only in isotropic media is \mathbf{P} parallel to \mathbf{E} and \mathbf{M} parallel to \mathbf{B} . Under time-dependent conditions, we must take into account, in addition, that any reaction of matter to an external solicitation takes some time. Consequently, we expect that the above equations should continue to hold as long as the fields change slowly enough, but not for faster rates of change.

Let us start by considering a dielectric medium. As we know, we can distinguish polar dielectrics, whose molecules have an intrinsic dipole moment, and non-polar dielectrics, whose molecules do not.

In the polar media, the intrinsic molecular dipoles that have chaotic directions in the absence of an applied field tend to orient in the direction of an applied field. The probability of being oriented in that direction increases with the field intensity. If the latter varies in time, the orientation process varies as a consequence, with some delay due to the inertia of the molecules. If the electric field varies too quickly, the molecules do not have time enough to reorient themselves, as they would under static or slowly varying conditions.

In the non-polar media, the polarization process is due to a rearrangement of the atomic structure under the action of the electric field. In Fig. 4.11, we pictured the atom as a sphere of negative charge (the electrons) with a point-like positive charge (the nucleus) in the center. The action of the applied field is to displace the negative charge center from the positive charge. This originates a restoring force proportional to their relative displacement. Being that the electron is quite light, compared to a molecule, they can rearrange quite rapidly, much more rapidly than the reorientation of the molecules, but some time is still needed.

We understand from these arguments that the value of the polarization $\mathbf{P}(t)$ at any given instant t does not only depend on the value of the electric field at that instant of time, but also on the values the field had at all the previous times, because matter is still reacting to those values. In other words, in a linear medium, $\mathbf{P}(t)$ depends linearly on the values of the field in all the instants $t - \tau$ (where $\tau > 0$ is the generic time interval in the past). The proportionality constant between polarization and field, which we call k , at the instant t shall depend on $\mathbf{E}(t - \tau)$ for all the $\tau > 0$. The polarization at the instant t is the sum, better expressed as the integral, of the contributions due to all the past instants. We can write that as

$$\mathbf{P}(t) = \int_0^{\infty} k(\tau) \mathbf{E}(t - \tau) d\tau. \quad (10.65)$$

Very similar arguments hold for paramagnetic and diamagnetic materials and a similar equation can be written for \mathbf{M} . These relations should be substituted for \mathbf{P} and \mathbf{M} in the Maxwell equations, which become quite difficult to express. Fortunately, however, the problem simplifies a lot when the fields depend on time

as circular functions, namely as $\cos \omega t$ (or as $\sin \omega t$, which is equivalent), whatever the angular frequency may be. Let us see how. Assume the electric field to be

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0(\mathbf{r}) \cos \omega t,$$

in which the space dependence given by the factor $\mathbf{E}_0(\mathbf{r})$ is arbitrary. To simplify the expressions, let us use complex formalism. The above expression is the real part of

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0(\mathbf{r}) e^{i\omega t}.$$

We can write Eq. (10.65) as

$$\mathbf{P}(\mathbf{r}, t) = \mathbf{E}_0(\mathbf{r}) \int_0^\infty k(\tau) e^{i\omega(t-\tau)} d\tau,$$

of which we should take the real part at the end. Now comes the magic, because we can write that as

$$\mathbf{P}(\mathbf{r}, t) = \mathbf{E}_0(\mathbf{r}) \int_0^\infty k(\tau) e^{i\omega t} e^{-i\omega\tau} d\tau = \mathbf{E}_0(\mathbf{r}) e^{i\omega t} \int_0^\infty k(\tau) e^{-i\omega\tau} d\tau,$$

where we could take the factor $e^{i\omega t}$, which is independent of the integration variable, out of the integral. This simplifies the issue enormously. As we see, the polarization at the instant t , namely $\mathbf{P}(\mathbf{r}, t)$, is proportional to the electric field at the *same* instant t . The proportionality constant is the integral on the right-hand side, namely $\int_0^\infty k(\tau) e^{-i\omega\tau} d\tau$, which is clearly time-independent. This “miracle” happens only if the fields depend on time as an exponential of the imaginary argument (of which cosine and sine are special cases, being its real and imaginary parts), because the exponential of the difference $t - \tau$ is the product of the exponentials of t and $-\tau$. In conclusion, in this case, the relations between \mathbf{P} and \mathbf{E} and between \mathbf{M} and \mathbf{B} valid under static conditions also hold for a $\cos \omega t$ dependence on time, with the difference that the proportionality constants (in practice χ_e and χ'_m) are functions of ω .

In conclusion, when the fields depend on time as

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0(\mathbf{r}) \cos \omega t, \quad \mathbf{B}(\mathbf{r}, t) = \mathbf{B}_0(\mathbf{r}) \cos \omega t. \quad (10.66)$$

the electric and magnetic susceptibilities are given by

$$\mathbf{P}(\mathbf{r}, t) = \epsilon_0 \chi_e(\omega) \mathbf{E}(\mathbf{r}, t), \quad \mathbf{M}(\mathbf{r}, t) = \frac{\chi'_m(\omega)}{\mu_0} \mathbf{B}(\mathbf{r}, t). \quad (10.67)$$

As already noticed, the values of both susceptibilities are equal or very close to their values under static conditions at low frequencies, namely small values of ω , where they can be considered to be constants. The minimum frequency at which a susceptibility becomes different from its static value depends on the material. It should already be clear to the reader that these values are smaller for polar dielectrics on one side and paramagnetics on the other. For example, the electric susceptibility of water already differs appreciably from its static value at about one megahertz if liquid, and as low as a few hundreds kilohertz if ice. At optical frequencies, which are on the order of hundreds of terahertz, the angular frequency dependence of the electric susceptibility is evident for all materials. Its effect is the *dispersion* of light, which originates phenomena like the rainbow. We shall see that in the 4th volume of this course.

In that 4th volume, we shall also study how functions with any “reasonable” dependence on time can be expressed as sums or integrals of terms of the type in Eq. (10.66). These are the Fourier series and the Fourier integral, after Jean Baptiste Fourier (France, 1768–1830). This property allows one to deal with arbitrary time-dependence of the fields.

10.8 Discontinuities of \mathbf{E} and \mathbf{B}

In Sect. 1.14, we saw that the electric field under static conditions has a discontinuity crossing a charged surface. More precisely, if Σ is such a surface and σ its surface charge density, the components of \mathbf{E} tangent to Σ are continuous, while the normal component has the discontinuity σ/ϵ_0 , namely

$$\Delta\mathbf{E}_T = 0, \quad \Delta E_N = \sigma/\epsilon_0. \quad (10.68)$$

Similarly, in Sect. 6.7, we saw that the magnetic field under steady conditions is discontinuous crossing a surface Σ carrying the surface current density \mathbf{k} . The field component tangent to Σ perpendicular to \mathbf{k} has the discontinuity $\mu_0 k$, while the field component normal to the surface is continuous, namely

$$\Delta B_T = \mu_0 k, \quad \Delta B_N = 0. \quad (10.69)$$

We shall now prove that these relations hold in general, namely also under time-dependent conditions.

The conclusion is immediately obvious for the normal component, both of \mathbf{E} and of \mathbf{B} , because the results were obtained starting from the equations of the divergence of the fields, which are the same under time-independent and time-dependent conditions.

The behavior of the tangent components depends, contrastingly, on the curl of the fields, which are different under static and dynamic conditions.

Fig. 10.14 A charge or current-carrying surface and the circuit Γ on which the Stokes theorem is applied

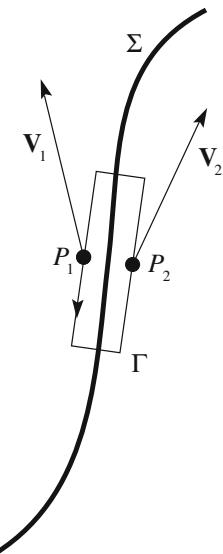


Figure 10.14 is drawn for both the electric and the magnetic cases, representing the surface Σ carrying the surface charge density σ in one case, the surface current density \mathbf{k} in the other. The latter vector is normal to the page. The vectors \mathbf{V}_1 and \mathbf{V}_2 are the field, \mathbf{E} in one case, \mathbf{B} in the other. The curve Γ is a parallelogram. The lengths dl of its sides parallel to Σ are infinitesimal, while those normal to Σ are infinitesimally of the higher order, in order to be able to neglect their contributions to the circulation integrals.

Let us start with the electric field. The circulation of \mathbf{E} about Γ is equal to the opposite of the rate of change of the flux of the magnetic field linked to Γ . Even if this quantity is not rigorously zero, as in statics, the magnetic flux and, consequently, its rate of change are proportional to the area inside Γ , which is infinitesimally of higher order than dl . Consequently, the circulation of \mathbf{E} about Γ is infinitesimally of superior order, and we can write $[E_T(P_2) - E_T(P_1)]dl = 0$. In conclusion, the tangent component of \mathbf{E} is continuous, as in statics.

In the case of the circulation of \mathbf{B} , a completely similar argument establishes that the contribution of the displacement current, the one that is not present in statics, is infinitesimally of superior order. Consequently, the first Eq. (10.69) also holds under dynamic conditions.

10.9 The Electromagnetic Potentials

In electrostatics, we introduced the electrostatic potential, which is a scalar function that we have called ϕ . Similarly, in magnetostatics, we introduced the vector potential \mathbf{A} . Both quantities have been defined for fields that do not depend on time.

We shall now generalize the definition to dynamic conditions. We shall use the term *scalar potential* for ϕ and keep the name of vector potential for \mathbf{A} .

In electrostatics, we could define ϕ to be such that the electric field is the opposite of its gradient, because the electric field curl is identically zero. This does not prove true under dynamic conditions, so we need to try to find an irrotational vector. On the other hand, we defined the vector potential to be such that

$$\nabla \times \mathbf{A} = \mathbf{B}. \quad (10.70)$$

This definition is allowed if the divergence of \mathbf{B} is zero. But this is true also under time-dependent conditions. We can then keep Eq. (10.70) as the definition of \mathbf{A} in general.

We now substitute into the Faraday law (Eq. 10.50) \mathbf{B} given by Eq. (10.70) and obtain

$$\nabla \times \mathbf{E} = -\frac{\partial \nabla \times \mathbf{A}}{\partial t},$$

which can obviously be written as

$$\nabla \times \left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} \right) = 0.$$

Well, $\mathbf{E} + \partial \mathbf{A} / \partial t$ on the left-hand side is an irrotational vector that reduces to \mathbf{E} under time-independent conditions. We can use it to define a scalar function ϕ , such as

$$\nabla \phi = -\mathbf{E} - \frac{\partial \mathbf{A}}{\partial t}. \quad (10.71)$$

The scalar potential defined in this way reduces to the electrostatic potential under static conditions, as it should.

Summarizing, we define the electromagnetic potentials as two fields, one scalar and one vector, such that

$$\mathbf{B} = \nabla \times \mathbf{A}, \quad \mathbf{E} = -\nabla \phi - \frac{\partial \mathbf{A}}{\partial t}. \quad (10.72)$$

We saw that, under time-independent conditions, the potentials are not completely defined. The same holds under the general conditions we are now considering. Indeed, any ϕ' and \mathbf{A}' that are such give the same electric and magnetic fields \mathbf{E} and \mathbf{B} through Eq. (10.72), as ϕ and \mathbf{A} are equivalent electromagnetic potentials.

A vector field \mathbf{A}' obtained by adding to \mathbf{A} the gradient of any scalar field ψ , namely $\mathbf{A} \rightarrow \mathbf{A}' = \mathbf{A} + \nabla \psi$, gives the same magnetic field, because $\nabla \times \mathbf{A} = \nabla \times \mathbf{A}'$. This is not enough, however, because in dynamics, the electric

field depends not only on ϕ , but on \mathbf{A} as well, as in Eq. (10.72). It is easy to show that neither \mathbf{E} nor \mathbf{B} change when we *contemporarily* do the substitutions

$$\mathbf{A} \rightarrow \mathbf{A}' = \mathbf{A} + \nabla\psi, \quad \phi \rightarrow \phi' = \phi - \frac{\partial\psi}{\partial t}. \quad (10.73)$$

Indeed, we then have, for the electric and magnetic fields, the substitutions

$$\begin{aligned}\mathbf{B} &\rightarrow \mathbf{B}' = \nabla \times \mathbf{A}' = \nabla \times \mathbf{A} = \mathbf{B} \\ \mathbf{E} &\rightarrow \mathbf{E}' = -\nabla\phi' - \frac{\partial\mathbf{A}'}{\partial t} = -\nabla\phi + \frac{\partial}{\partial t}\nabla\psi - \frac{\partial\mathbf{A}}{\partial t} - \frac{\partial}{\partial t}\nabla\psi = \mathbf{E}.\end{aligned}$$

The electric and magnetic fields do not change. All the observable effects depend on the forces acting on the charges, which are due to the electric and magnetic fields, not directly to the electromagnetic potentials. We are then free to choose any electromagnetic potential satisfying Eq. (10.72). This is the gauge-fixing freedom that we already exploited in magnetostatics, in Sect. 6.8. We phrase this conclusion by stating that electromagnetism (namely the Maxwell equations) is (are) invariant under the gauge transformations of the potentials in Eq. (10.73). This property is called gauge invariance. We observe here that gauge invariance is, in classical electromagnetism, just a property that turns out to be very useful for simplifying equations, as we shall do immediately, but no more than that. In quantum electrodynamics, and more generally, in the quantum theories of all the fundamental interactions (gravity excluded), gauge invariance assumes an extremely fundamental role.

Coming back to electromagnetic potentials, we can say that Eq. (10.72) allows us to calculate the electric and magnetic fields once we know the electromagnetic potentials. Consequently, we can describe the electromagnetism in terms of the potentials, rather than of the fields. Let us find the partial differential equations ruling the potentials, which are obviously equivalent to the Maxwell equations.

We have already used two Maxwell equations in the definitions of the potentials, namely the equations of the divergence of \mathbf{B} and of the curl of \mathbf{E} . Let us now substitute to \mathbf{E} in Eq. (10.49) its expression Eq. (10.72). We obtain

$$\nabla \cdot \left(-\nabla\phi - \frac{\partial\mathbf{A}}{\partial t} \right) = \frac{\rho}{\epsilon_0},$$

which we can write as

$$\nabla^2\phi + \frac{\partial\nabla \cdot \mathbf{A}}{\partial t} = -\frac{\rho}{\epsilon_0}. \quad (10.74)$$

Unfortunately, it contains both \mathbf{A} and ϕ . We still have not used the fourth Maxwell equation. Let us substitute in Eq. (10.52) the expressions Eq. (10.72). We have

$$\nabla \times (\nabla \times \mathbf{A}) - \mu_0 \epsilon_0 \frac{\partial}{\partial t} \left(-\nabla \phi - \frac{\partial \mathbf{A}}{\partial t} \right) = \mu_0 \mathbf{j}.$$

We can simplify it a bit by using the vector identity

$$\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}.$$

We obtain, after a few rearrangements,

$$\nabla^2 \mathbf{A} - \mu_0 \epsilon_0 \frac{\partial^2 \mathbf{A}}{\partial t^2} - \nabla \cdot \left(\nabla \cdot \mathbf{A} + \mu_0 \epsilon_0 \frac{\partial \phi}{\partial t} \right) = \mu_0 \mathbf{j}. \quad (10.75)$$

This equation is still quite complicated, but we can make it simple by exploiting the gauge invariance with a convenient choice of the divergence of \mathbf{A} . We made something similar in Sect. 6.8, where we simplified the equations choosing $\nabla \cdot \mathbf{A} = 0$, which is called the Coulomb gauge. In the present case, the convenient choice is

$$\nabla \cdot \mathbf{A} = -\mu_0 \epsilon_0 \frac{\partial \phi}{\partial t}, \quad (10.76)$$

which is called the Lorenz gauge, from Ludwig Lorenz (Denmark, 1829–1891). With this gauge, the terms in parenthesis on the left-hand side of Eq. (10.75) cancel one another out, and the equation becomes

$$\nabla^2 \mathbf{A} - \mu_0 \epsilon_0 \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu_0 \mathbf{j}, \quad (10.77)$$

which is now a quite simple equation in the vector potential alone. But there is a further benefit. Indeed, Eq. (10.74) becomes

$$\nabla^2 \phi - \mu_0 \epsilon_0 \frac{\partial^2 \phi}{\partial t^2} = -\frac{\rho}{\epsilon_0}. \quad (10.78)$$

which not only contains ϕ only, but is the same partial differential equation!

We have found four relations in total (three for the components of the vector potential and one for the scalar potentials), which are the same non-homogeneous partial differential equations, with, on the right-hand sides, the components of the current density and the charge density. We can write them with the speed of light in place of $\epsilon_0 \mu_0$ as

$$\begin{aligned} \nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} &= -\mu_0 \mathbf{j} \\ \nabla^2 \phi - \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} &= -\frac{\rho}{\epsilon_0}. \end{aligned} \quad (10.79)$$

with the gauge condition

$$\nabla \cdot \mathbf{A} + \frac{1}{c^2} \frac{\partial \phi}{\partial t} = 0. \quad (10.80)$$

The simplicity (and elegance) of the potential equations immediately makes evident two fundamental properties of electromagnetism, namely the existence of electromagnetic waves propagating with the speed of light, as we shall immediately see, and the covariance of the electromagnetic equation under the Lorentz transformations, as we shall see in the next section.

We have already found, in Sect. 10.3, that the electric and magnetic fields in a vacuum obey a partial differential equation, which is the wave equation. The same is true for the potentials. Indeed, in a vacuum, where there are no charges and no currents, Eq. (10.79) become the homogeneous equations

$$\begin{aligned} \nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} &= 0 \\ \nabla^2 \phi - \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} &= 0, \end{aligned} \quad (10.82)$$

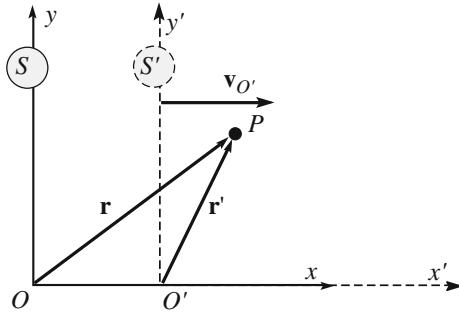
which, in both cases, are just the wave equation. The wave velocity is the inverse square root of the constant quantity multiplying the time second partial derivative, namely, as we well know, c . In the 4th volume of the course, we shall discuss wave phenomena in general and electromagnetic waves in particular. The electromagnetic waves in the wavelength, ranging between about 400 and 700 μm , can be seen by our own eyes. They are light.

10.10 Covariance of Electromagnetism

We have already recalled, in Sect. 6.16, that the relativity principle, which was established by Galileo Galilei in the XVIth century, states, in modern terms, that physical laws are covariant under transformations between two inertial reference frames in the relative motion of uniform translation. As discussed in the 1st volume, only two sets of transformation equations of the space and time coordinates between such frames exist: the Galilei transformations and the Lorentz transformations. The former are the limit of the latter for velocities much smaller than the speed of light. In the Galilei transformations, time intervals and distances between two points are invariant under the transformation, namely are the same for two observers in relative motion, while in the Lorentz transformations, they are different, but the velocity of light is invariant.

Let us briefly summarize what we discussed in the 1st volume. Figure 10.15 shows two inertial reference frames. The first one, which we call $S(x, y, z, t)$, has the coordinates x, y, z and time t . The second frame, $S'(x', y', z', t')$, has axes parallel

Fig. 10.15 Two inertial frames in relative uniform translator motion



to the first one. The relative velocity is along the x and x' axes that overlap. The constant velocity of S' , the velocity of its origin, is $v_{O'}$, and is in the positive direction of x . We choose the origins of the times in both frames at the instant in which O' and O coincide.

An *event* is something happening in a definite position and at a definite instant in time, as measured in the considered frame. It is a point in the space-time, which is a four-dimensional space. Following Henri Poincaré (France, 1854–1912), we can call the coordinates in the space-time

$$x_1 = x, \quad x_2 = y, \quad x_3 = z, \quad x_4 = ict \quad (10.83)$$

These are also the components of the basic four-vector in the space-time. Two are the relevant parameters in the Lorentz transformations. Both are pure numbers, which are functions of the velocity $v_{O'}$ of S' relative to S . The first one is the ratio of this velocity and the speed of light

$$\beta_{O'} = v_{O'}/c \quad (10.84)$$

the second is

$$\gamma_{O'} = 1/\sqrt{1 - \beta_{O'}^2} \quad (10.85)$$

The Lorentz transformations can be written in the form

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} \gamma & 0 & 0 & i\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -i\beta\gamma & 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \\ x'_4 \end{pmatrix} \quad (10.86)$$

The Lorentz transformations are rigid rotations in the space-time. Similarly to rigid rotation in the three-dimensional space, they leave the norms of the four-vectors invariant.

Maxwell equations are covariant under the Lorentz transformations, but not under the Galilei transformations. A consequence of that fact is the *invariance* of the speed of light in a vacuum. Let us consider the following two events in S . The first one is the start of a light pulse from its origin O at the instant $t = 0$, the second is the arrival of the pulse at the point (x, y, z) at time t . We express the fact that the speed of light is c , writing

$$x^2 + y^2 + z^2 - (ct)^2 = \sum_{j=1}^4 x_j^2 = 0$$

In S' , the norm of the four-vector has the same value, namely

$$\sum_{j=1}^4 x'_j^2 = x'^2 + y'^2 + z'^2 - (ct')^2 = 0$$

The statement we have just made, that Maxwell equations are covariant under Lorentz transformations, means that, if we transform the physical quantities present in the Maxwell equations (\mathbf{E} , \mathbf{B} , \mathbf{j} and ρ) with the Lorentz transformations, both sides of each equation change in the same way. Consequently, each equation that is valid in the first frame is also valid in the second.

We have seen in the 1st volume that some three-vectors (namely vectors in the three-dimensional space), but not all of them, can be “promoted” to four-vectors by including a fourth component. This is the case for the linear momentum \mathbf{p} , which forms, together with energy U , the four-momentum $(\mathbf{p}, iU/c)$. Similarly, we saw, in Sect. 6.16, that the three-vector current density \mathbf{j} can be “promoted” to four-vector with a fourth component proportional to the charge density, namely (\mathbf{j}, ipc) . This is called the four-current density. Consequently, we already know how these elements of the Maxwell equation transform. We do not, however, know the transformation equations for the electric field \mathbf{E} and the magnetic field \mathbf{B} . The “promotion” procedure does not work for these three-vectors, because the three components of each of them are not the space components of a four-vector. On the contrary, the six of them are the elements of a double antisymmetric tensor in four dimensions. Tensors are mathematical entities. We can think here of an antisymmetric 4×4 matrix, which has six independent elements. These are the electric and magnetic field components, factors c apart. We shall not find the Lorentz transformations for this tensor here, because there is a simpler way. Namely, we shall consider the covariance of the equations for the potentials, Eq. (10.79), which are equivalent to the Maxwell equations.

The right-hand side of the first Eq. (10.79) is proportional through μ_0 , which is an invariant constant, to the three-vector \mathbf{j} . The right-hand side of the second equation is almost, but not exactly, the fourth component of the four-current (\mathbf{j}, ipc) . In order to obtain it, we multiply the left-hand and right-hand sides by the two sides of the identity $i/c = ic\mu_0\varepsilon_0$, respectively and we re-write the equations in the form

$$\begin{aligned} \left[\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right] \mathbf{A} &= -\mu_0 \mathbf{j} \\ \left[\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right] i\phi/c &= -\mu_0(i\rho c) \end{aligned} \quad (10.87)$$

In this form, the equations, which are, in fact, the same equation, are very similar to the Laplace equation, which we found under time-independent conditions for the electrostatic potential, Eq. (1.87), and for the vector potential, Eq. (6.43). The operator on the left-hand side of the Laplace equation is the Laplacian

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

This, we recall, is the norm of the gradient, which is the three-vector operator

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right)$$

In place of the Laplacian, we now have

$$\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2},$$

which we can write as

$$\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} + \frac{\partial^2}{\partial(ict)^2} = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} + \frac{\partial^2}{\partial x_4^2}.$$

We immediately recognize this expression as the norm of the four-dimensional gradient, which we write as

$$\partial_\mu \equiv \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}, \frac{\partial}{\partial x_4} \right) = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}, \frac{\partial}{\partial ict} \right). \quad (10.88)$$

The norm of any three-dimensional vector, such as the three-dimensional gradient, is invariant under rotations of the axes, namely it is a three-scalar. In exactly the same way, the norm of any four-dimensional vector, such as the four-dimensional gradient, is invariant under Lorentz transformations, namely it is a four-scalar. We write this operator, which is called the d'Alembert operator or d'Alembertian, after Jan Baptiste d'Alembert (France, 1717–1783), as

$$\square^2 \equiv \sum_{\mu=1}^4 \frac{\partial^2}{\partial x_\mu^2} = \left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right). \quad (10.89)$$

Equation (10.87) are

$$\square^2 \mathbf{A} = -\mu_0 \mathbf{j}, \quad \square^2 i\phi/c = -\mu_0(i\rho c). \quad (10.90)$$

Now, we have on the right-hand sides the components of a four-vector, on the left-hand side a scalar operator times four quantities, which should consequently be the components of a four-vector as well. This statement can be rigorously proven, but we shall skip that. The four-vector is called the four-potential

$$A_\mu = (\mathbf{A}, i\phi/c). \quad (10.91)$$

We still have one equation to consider, namely the gauge condition (Eq. 10.80). The relativity principle requires that this equation should be covariant as well. We start by observing that the first term is the divergence of the vector potential. Let us then look at the divergence in four dimensions of the four-potential. This is

$$\sum_{\mu=1}^4 \frac{\partial A_\mu}{\partial x_\mu} = \nabla \cdot \mathbf{A} + \frac{\partial(i\phi/c)}{\partial(ict)} = \nabla \cdot \mathbf{A} + \frac{1}{c^2} \frac{\partial \phi}{\partial t},$$

which is just the left-hand side of Eq. (10.80), which can then be written as

$$\sum_{\mu=1}^4 \frac{\partial A_\mu}{\partial x_\mu} = 0. \quad (10.92)$$

The four-divergence being a four-scalar, the equation is invariant.

We have thus shown that the electromagnetic equations obey the relativity principle, provided the Lorentz transformations are used.

10.11 Physical Meaning of the Electromagnetic Potentials

As we have discussed, the electromagnetic field can be mathematically described either in terms of the fields \mathbf{E} and \mathbf{B} , or in terms of the potentials ϕ and \mathbf{A} . While the two descriptions are completely mathematically equivalent, they are not necessarily so physically. In this section, we shall discuss the following questions. Are the potentials measurable, or, in other words, can we define them operationally? What is the meaning of the potentials as physical (rather than mathematical) objects?

We recall the operational definitions of the fields \mathbf{E} and \mathbf{B} . We considered an exploring charge q at the space point \mathbf{r} and time instant t in which the fields had to be defined. We measured the force acting on the charge. If the charge is at rest, we measure the force $\mathbf{F}(\mathbf{r}, t) = q\mathbf{E}(\mathbf{r}, t)$ and, q being known, we determine $\mathbf{E}(\mathbf{r}, t)$. If the charge moves with known velocity \mathbf{v} , we can determine $\mathbf{B}(\mathbf{r}, t)$ by measuring the force $\mathbf{F}(\mathbf{r}, t) = q\mathbf{E}(\mathbf{r}, t) + q\mathbf{v} \times \mathbf{B}(\mathbf{r}, t)$. Note that, the magnetic force being $q\mathbf{v} \times \mathbf{B}(\mathbf{r}, t)$ in any

case perpendicular to velocity, we must measure one component of \mathbf{B} at a time, for different directions of the velocity of the test charge.

We think of fields \mathbf{E} and \mathbf{B} as *real* physical quantities, mainly because we need them to get rid of actions at a distance. Indeed, in the electric case, we may describe, for example, the action of the point charge q_1 at point \mathbf{r}_1 on point charge q_2 at another point \mathbf{r}_2 by the force exerted by the former on the latter (the Coulomb force if the charges are at rest). In this description, we have an action at a distance. We have logically gotten rid of that by thinking that the charge q_1 produces an electric field \mathbf{E} in all the surrounding space independently of the presence of q_2 . When q_2 is present in \mathbf{r}_2 , it is subject to a force given by the pre-existent field in \mathbf{r}_2 times q_2 . The latter description is experimentally verified under time-dependent conditions, when we observe that the interaction propagates with a finite velocity, which is the velocity of light. Completely similar arguments hold for the magnetic field \mathbf{B} .

On the other hand, we need neither the scalar potential ϕ nor the vector potential \mathbf{A} to avoid actions logically at distance, because all observable effects are the result of the forces. The force on the generic charge q at the space point \mathbf{r} in the time instant t moving with velocity \mathbf{v} depends on the fields at that point and in that instant, as $\mathbf{F}(r, t) = q[\mathbf{E}(r, t) + \mathbf{v} \times \mathbf{B}(r, t)]$. Consequently, in a region of space in which the fields are zero and the potentials are different from zero, no force acts on the charge and no effects are observable. From this point of view, the potentials look to be useful mathematical objects deprived of a substantial physical meaning. However, this conclusion is premature.

Let us start with the operational definition of the scalar potential, or, in other words, with its measurability. The scalar potential ϕ can be measured (within an additive constant, which is its gauge freedom) if the electric field is constant with time during the measurement operation. In this case, the definition of ϕ is the definition known in electrostatics. Namely, the potential $\phi(\mathbf{r})$ at point \mathbf{r} is the potential energy of the unit charge at that point. Let the charge q move under the action of the field from \mathbf{r}_1 to \mathbf{r}_2 . The sum of potential energy ($q\phi$) and kinetic energy (U_K) of the charge is constant. We determine the potential difference between those points, namely $\phi(\mathbf{r}_2) - \phi(\mathbf{r}_1)$, by measuring the kinetic energies of the charge U_{K1} when it is in \mathbf{r}_1 and U_{K2} when it is in \mathbf{r}_2 . Note that this operation must be done without inducing any change in the electric field. Namely, the charge q should be so small that the positions of the charge sources of the field do not change when it moves. Energy conservation then gives us

$$q\phi(\mathbf{r}_1) + U_{K1} = q\phi(\mathbf{r}_2) + U_{K2},$$

and consequently

$$\phi(\mathbf{r}_2) - \phi(\mathbf{r}_1) = (U_{K2} - U_{K1})/q. \quad (10.93)$$

We repeat that the equations we just wrote are true only if the electric field remains constant during the measurement operations. Under these conditions, we can define the scalar potential operationally, up to an additive constant.

Beyond being measurable, the scalar potential has the following physical meaning. If we want to know the energy of the charge q in \mathbf{r} using only local information, then we must use ϕ , which thus gets a physical meaning. Indeed, $\phi(\mathbf{r})$ is the interaction energy of the unit charge in \mathbf{r} . In other words, $q\phi(\mathbf{r})$ is the fraction of the field energy that is *locally* available to be transformed into kinetic energy of the charge. Still, in other words, $q\phi(\mathbf{r})$ is the work to be done against the field forces to move the charge q from an infinite distance, where it does not interact with the field, to the position \mathbf{r} . Notice again that the operation must be done without affecting the positions of the charge sources of the potential $\phi(\mathbf{r})$. When the charge is positioned in \mathbf{r} , the energy of the system is changed by $q\phi(\mathbf{r})$. At the end of this section, we shall prove this statement with an explicit calculation.

Let us now look at the operational definition of the vector potential \mathbf{A} , namely at its measurability. We must measure one of the components of \mathbf{A} at a time, much as we did for the measurement for the magnetic field. We have just seen that the scalar potential can be measured under the conditions of having the electric field constant in time. An analogous condition holds for the measurability of the components of \mathbf{A} . In this case, in place of energy conservation, we start from the equation of motion of the charged particle and write it in terms of the potentials instead of the fields, using Eq. (10.72). If \mathbf{p} is its momentum and \mathbf{v} the velocity of the charge, we have

$$\frac{d\mathbf{p}}{dt} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) = q\left(-\nabla\phi - \frac{\partial\mathbf{A}}{\partial t} + \mathbf{v} \times \nabla \times \mathbf{A}\right), \quad (10.94)$$

which can be written as

$$\frac{\partial}{\partial t}(\mathbf{p} + q\mathbf{A}) = -q(\nabla\phi - \mathbf{v} \times \nabla \times \mathbf{A}).$$

On the right-hand side, we have the gradient of the scalar potential and the cross-product of velocity and the curl of \mathbf{A} . The latter can be transformed into another gradient through some fairly tricky vector algebra, which we skip, and by benefit of the gauge freedom choosing the Coulomb gauge, namely $\nabla \cdot \mathbf{A} = 0$. The result is $\mathbf{v} \times \nabla \times \mathbf{A} = \nabla(\mathbf{v} \cdot \mathbf{A})$. The above equation becomes

$$\frac{\partial}{\partial t}(\mathbf{p} + q\mathbf{A}) = -\nabla[q(\phi - \mathbf{v} \cdot \mathbf{A})], \quad (10.95)$$

where the gradient only acts on the potentials, not on velocity. We can measure a component of the vector potential if the time derivative of that component on the left-hand side is zero. To guarantee that, we operate with that component of the gradient on the right-hand side equal to zero.

Figure 10.16 shows an infinitely long solenoid carrying a steady current of intensity I . Outside the solenoid, the electric and magnetic fields are zero. The first term on the right hand side of Eq. (10.95), which is proportional to the gradient of the scalar potential, is hence zero. We profit of the geometry of the problem and use

cylindrical polar coordinates, with z on the symmetry axis, r as the distance from the axis and α as the azimuth angle. As we know, the lines of the vector potential are circumferences normal to the axis with centers on it. The only non-zero component of the vector potential is the azimuthal component A_α , which we now want to measure. This component has constant values at all the points of one of these circumferences, like Γ in Fig. 10.16. Correspondingly, $\nabla(\mathbf{v} \cdot \mathbf{A}) = 0$ along Γ . We then position a rigid plastic wire about Γ and pass it through the hole that we have done a small bead. We place the charge q on the bead, which can move on the circumference with negligible friction. In this way, we prepared the system constraining the charge to move at points at which the right-hand side of Eq. (10.95) is zero, and we can do the measurement.

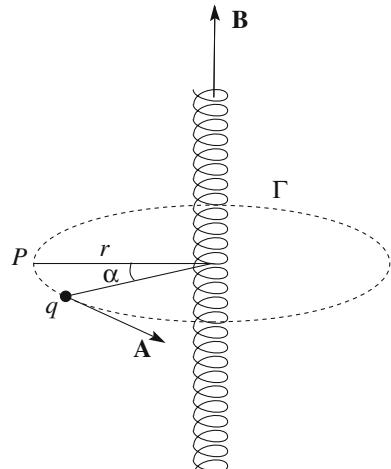
We now have to take the current intensity slowly down to zero. The flux of \mathbf{B} through Γ decreases and an emf is induced about Γ . This is exactly the effect considered in Sect. 10.6, when we discussed the Feynman paradox. We can say, in an equivalent manner, that an electric field \mathbf{E} is induced at the points of Γ as a consequence of the rate of change of A_α . \mathbf{E} is tangent to Γ , namely its only component is $E_\alpha = -\partial A_\alpha / \partial t$, for Eq. (10.72). This field acts on q with the force qE_α , which, in every infinitesimal time interval dt , changes the p_α component of the momentum of the charge by

$$dp_\alpha = qE_\alpha dt = -q \frac{\partial A_\alpha}{\partial t} dt = -q dA_\alpha,$$

which can be written as

$$d(p_\alpha + qdA_\alpha) = 0. \quad (10.96)$$

Fig. 10.16 A very long solenoid and a charge moving in a region where $\mathbf{B} = 0$ and $\mathbf{A} \neq 0$



This means that the quantity that we call

$$\Pi_\alpha \equiv p_\alpha + qA_\alpha. \quad (10.97)$$

is a constant of the motion of the charge.

Π_α is the α -component of the vector Π , which is called the *canonical momentum*, namely

$$\Pi = \mathbf{p} + q\mathbf{A}. \quad (10.98)$$

Equation (10.97) shows that the α component of the canonical momentum of the system is constant in time.

In the initial state of our experiment the charge q is at rest in a location in which the magnetic field is zero but A_α is different from zero, and in the final state, $A_\alpha = 0$ and the charge has a non-zero linear momentum, which we call $p_{\alpha f}$. We can find this quantity using the canonical momentum conservation (of its α -component, to be precise). We write

$$\Pi_{\alpha, \text{iniz}} = 0 + qA_\alpha = \Pi_{\alpha, \text{fin}} = p_{\alpha f} + 0,$$

namely

$$p_{\alpha f} = qA_\alpha. \quad (10.99)$$

Hence, when the field is switched off, the charged particle acquires a linear momentum equal to its charge q times the initial vector potential. We can say that $q\mathbf{A}$ is the *potential linear momentum* of q in the field, namely the field linear momentum that is locally available to be transformed into linear momentum of matter. We can say that the potential action of \mathbf{A} is local, while that of \mathbf{B} is at a distance. Still, in other words, moving the charge q from infinite distance to the point P in Fig. 10.16, the linear momentum of the system (field plus particle) changes by $q\mathbf{A}$, which is then the interaction momentum between charge and field. Once more, the operation must be done without affecting the system of currents producing the magnetic field. We shall check this statement with an explicit calculation at the end of the section.

Historically, the meaning of the vector potential as the potential momentum stored in the field was already recognized by James Clerk Maxwell. He wrote in “A treatise on electricity and magnetism” in 1873:

The vector \mathbf{A} represents in direction and magnitude the time-integral of the electromotive force which a particle placed at the point (x, y, z) would experience (*gaining an equal momentum*) if the primary current were suddenly stopped. We shall therefore call it Electrokinetic Momentum (*term no longer used*) at the point (x, y, z) .

In 1903, Joseph John Thomson, in his lectures at Yale University, published in the book “Electricity and Matter” the following year, discussed this issue through several examples. He wrote:

It is important to bear in mind that this momentum is not in any way different from ordinary mechanical momentum and can be given up or taken from the momentum of moving bodies.

and, concluding after his demonstrations:

Thus, whether the magnetic field is due to permanent magnets or to electric currents or partly to one and partly to the other, the momentum when an electrified point is placed at P is equivalent to a momentum $q\mathbf{A}$ at P where \mathbf{A} is the vector potential at P .

Note that the scalar potential ϕ is *interaction energy* per unit charge, and is linked to the electric field, which modifies the kinetic energy of the charges. On the other hand, the vector potential \mathbf{A} is *potential linear momentum* and is linked to the magnetic field, which modifies the direction of the momentum of the charged particles and not their energy.

Let us now analyze how the gauge invariance property appears in the experiment we are considering. The observable, namely the change in the momentum of the charge q , is due to the induced emf

$$\oint_{\Gamma} \mathbf{E} \cdot d\mathbf{s} = -\frac{d}{dt} \oint_{\Gamma} \mathbf{A} \cdot d\mathbf{s} = -\frac{d}{dt} \int_{\Sigma} (\nabla \times \mathbf{A}) \cdot \mathbf{n} d\Sigma,$$

where, on the right-hand side, we have used the Stokes theorem on a surface Σ bound by Γ . The equation is obviously the Faraday law. We wrote it down to observe that the momentum acquired by q is determined by the curl of \mathbf{A} and is independent of its divergence. Consequently, the experiment cannot distinguish between two vector potentials \mathbf{A} and \mathbf{A}' having identical curl and different divergence. This is clearly the gauge invariance.

We conclude by stating that it is possible to define each component of the vector potential operationally, provided we operate keeping the homologous component of the canonical momentum, constant according to Eq. (10.95).

We observe here that, in quantum physics, the relevance of the electromagnetic potentials is absolutely fundamental. Indeed, in quantum physics, the forces are not very relevant, while energy and momentum become of central importance. As a matter of fact, waves are associated with quantum systems. The frequency of the wave is proportional to energy (to the total energy, namely interaction plus kinetic energy) and the wavelength is inversely proportional to momentum (to the canonical momentum, namely mechanical plus interaction momentum). The most natural description is in terms of the potentials rather than of the electric and magnetic fields. Observable effects exist when, for example, a beam of charged particles crosses a region of zero electric and magnetic fields and non-zero potentials, such as outside a current-carrying solenoid, as shown in Fig. 10.16. The effect is known as the Aharonov-Bohm solenoid effect, after Yakir Aharonov (Israel, 1932-) and David Bohm (USA, 1917–1992), who predicted it in 1959. The effect was later observed experimentally. The vector potential plays crucial roles in superconductivity and quantum field theories.

Let us now calculate the interaction energy and the interaction momentum between a point-like particle of charge q , at rest in \mathbf{r}_1 , and a time-independent electromagnetic field $\mathbf{E}_0, \mathbf{B}_0$. The total electric field is the sum of \mathbf{E}_0 and the electric field of charge q , which we call \mathbf{E}_q . The total magnetic field is just \mathbf{B}_0 because, the charge being at rest, its magnetic field is zero. Its electric field at the generic point \mathbf{r}_2 is

$$\mathbf{E}_q(\mathbf{r}_2) = \frac{q}{4\pi\epsilon_0 r_{12}^2} \frac{\mathbf{u}_{12}}{r_{12}^2}. \quad (10.100)$$

We obtain the energy of the total field by integrating the energy density given by Eq. (10.36), namely

$$\begin{aligned} U &= \frac{\epsilon_0}{2} \int (\mathbf{E}_0 + \mathbf{E}_q)^2 dV + \frac{1}{2\mu_0} \int \mathbf{B}_0^2 dV \\ &= \left[\frac{\epsilon_0}{2} \int \mathbf{E}_0^2 dV + \frac{1}{2\mu_0} \int \mathbf{B}_0^2 dV \right] + \frac{\epsilon_0}{2} \int \mathbf{E}_q^2 dV + \epsilon_0 \int \mathbf{E}_0 \cdot \mathbf{E}_q dV. \end{aligned}$$

Let us look at the right-hand side of this equation. The first term, in brackets, is the energy of the field $\mathbf{E}_0, \mathbf{B}_0$ alone, the second term is the field of the point charge alone. These two terms also exist if the charge is outside the field. The third term is the interaction energy between the charge and the $\mathbf{E}_0, \mathbf{B}_0$ field, when the charge is at \mathbf{r}_1 . Indicating with ϕ_q the potential of the charge, we can write

$$U_{\text{int}} = \epsilon_0 \int \mathbf{E}_0(\mathbf{r}_2) \cdot \mathbf{E}_q(\mathbf{r}_2) dV_2 = -\epsilon_0 \int \mathbf{E}_0(\mathbf{r}_2) \cdot \nabla \phi_q(\mathbf{r}_2) dV_2.$$

We now use the identity

$$\mathbf{E} \cdot \nabla \phi = \nabla(\mathbf{E}\phi) - \phi \nabla \cdot \mathbf{E}$$

and write

$$U_{\text{int}} = -\epsilon_0 \int_{\text{allspace}} \nabla \cdot [\mathbf{E}_0(\mathbf{r}_2) \phi_q(\mathbf{r}_2)] dV_2 + \epsilon_0 \int_{\text{allspace}} \phi_q(\mathbf{r}_2) \nabla \cdot \mathbf{E}_0(\mathbf{r}_2) dV_2.$$

The first integral on the right-hand side is the integral to the entire space of a divergence. It can be shown to be zero in the hypothesis that the charge distribution originating the field does not extend to infinite. The procedure is the same one that we used when we calculated the electric field energy and the magnetic field energy. Namely, we start with an integral of the divergence over a sphere of large radius, so large as to contain all the charges. Then, we use the Gauss divergence theorem to convert the integral into the integral of a flux over the surrounding surface. Then, we send the radius R of the surface to infinity. The integrand tends to zero as $1/R^3$, while the area diverges as R^2 . Consequently, the integral vanishes as $1/R$, and we are left with

$$U_{\text{int}} = \varepsilon_0 \int_{\text{allspace}} \phi_q(\mathbf{r}_2) \nabla \cdot \mathbf{E}_0(\mathbf{r}_2) dV_2,$$

in which we substitute

$$\phi_q(\mathbf{r}_2) = \frac{1}{4\pi\varepsilon_0} \frac{q}{r_{12}}$$

and

$$\nabla \cdot \mathbf{E}_0 = \rho/\varepsilon_0$$

where ρ is the charge density producing \mathbf{E}_0 . We obtain

$$U_{\text{int}} = \frac{q}{4\pi\varepsilon_0} \int \frac{\rho(\mathbf{r}_2)}{r_{12}} dV_2,$$

which we immediately recognize as

$$U_{\text{int}} = q\phi(\mathbf{r}_1). \quad (10.101)$$

That is what we wanted to show. The interaction energy between point charge and field is simply given by the well-known expression of the potential energy of the charge in the given static field. If the charge is in a region of zero \mathbf{E}_0 field, it interacts with that field anyway, in the sense that there is an interaction energy (corresponding to the work done when assembling the system). Indeed, the field of the charge is not confined to the site of the charge, but rather it invades all space, including where \mathbf{E}_0 is not zero. Together, they give origin to the interaction energy.

Let us now calculate the interaction momentum, by integration of the linear momentum density given by Eq. (10.40), namely

$$\mathbf{P} = \varepsilon_0 \int \mathbf{E} \times \mathbf{B} dV. \quad (10.102)$$

We obtain

$$\mathbf{P} = \varepsilon_0 \int \mathbf{E}_0(\mathbf{r}_2) \times \mathbf{B}_0(\mathbf{r}_2) dV_2 + \varepsilon_0 \int \mathbf{E}_q(\mathbf{r}_2) \times \mathbf{B}_0(\mathbf{r}_2) dV_2.$$

The first term on the right-hand side is the momentum of the given (stationary) field. The momentum of the field of the charge, which is at rest, is zero, because the magnetic field is zero. The second term on the right-hand side is the momentum due to the interaction between charge and $\mathbf{E}_0, \mathbf{B}_0$ field, and we write

$$\mathbf{P}_{\text{int}} = \varepsilon_0 \int \mathbf{E}_q(\mathbf{r}_2) \times \mathbf{B}_0(\mathbf{r}_2) dV_2 = -\varepsilon_0 \int \nabla \phi_q(\mathbf{r}_2) \times \mathbf{B}_0(\mathbf{r}_2) dV_2.$$

The integral on the right-hand side can be calculated by parts. Once more, the integrated part, which is a divergence, vanishes at infinite, and we are left with

$$\mathbf{P}_{\text{int}} = \varepsilon_0 \int \phi_q(\mathbf{r}_2) \nabla \times \mathbf{B}_0(\mathbf{r}_2) dV_2.$$

Now, we substitute the above expression for ϕ_q and $\nabla \times \mathbf{B} = \mu_0 \mathbf{j}$, where \mathbf{j} is the (steady) current density generating \mathbf{B}_0 , obtaining

$$\mathbf{P}_{\text{int}} = \frac{q\mu_0}{4\pi} \int \frac{\mathbf{j}(\mathbf{r}_2)}{r_{12}} dV_2,$$

which we recognize as

$$\mathbf{P}_{\text{int}} = q\mathbf{A}(\mathbf{r}_1), \quad (10.103)$$

which is what we wanted to show. The interaction momentum is also different from zero when the charge q is outside the solenoid, where the magnetic field is zero, but the vector potential is not. Indeed, the electric field \mathbf{E}_q of the charge extends to all space, even far from the position of the charge, and penetrates inside the solenoid, where, together with the magnetic field \mathbf{B}_0 , it originates the interaction momentum.

Summary

In this chapter, we completed the equations ruling the electric and magnetic fields under general conditions, in both their differential and integral forms. We learned the following principal concepts:

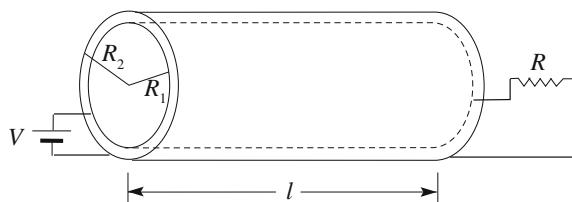
1. The four Maxwell equations
2. The displacement current.
3. Electromagnetic radiation.
4. The existence of electromagnetic waves and their speed in vacuum. Light is an electromagnetic wave.
5. Energy density and energy flux in an electromagnetic field.
6. Linear momentum density and flux in an electromagnetic field.
7. Matter and field linear momenta do not conserve separately but their sum does, if the system is isolated. The same holds for angular momenta.
8. The action and reaction law does not hold for a system of moving charges.
9. The electromagnetic potentials.
10. Gauge invariance of electromagnetism.

11. Covariance of electromagnetism under Lorentz transformations.
12. Physical meaning of the electromagnetic potentials.

Problems

- 10.1. Can the lines of the electric field be closed? If yes, what is the origin of the field? Can those lines exit from a point? If yes, what is the origin of the field? Answer the same questions for the magnetic field.
- 10.2. Could you measure the divergence of the vector potential?
- 10.3. A cylindrical wire conductor has length l and resistance per unit length r . We apply a potential difference V to its extremes. Calculate the Poynting vector immediately outside the surface of the conductor and the energy flux. Compare with the heat dissipated by the Joule effect.
- 10.4. Consider a solenoid made of a tightly packed cylindrical helix of a resistive wire and carrying a steady current I . Draw a sketch at a point immediately outside the solenoid with the following vectors: electric field, magnetic field, Poynting vector.
- 10.5. Are the discontinuities of the electric field components across a charged surface different under dynamic conditions compared with static ones? And the discontinuities of the magnetic field components across a current-carrying surface?
- 10.6. The distance between the metallic circular plates of a parallel plate capacitor is much smaller than their radius. The capacitor, which is initially charged, is being discharging on a resistor. What are direction and sense of the energy flow at a point between the plates close to the rim?
- 10.7. In a certain space region, the scalar potential is uniform and constant in time. The space and time dependence of the vector potential is $\mathbf{A}(x, y, z, t) = \mathbf{A}_0 + \mathbf{k}(x, y, z)\cos(\omega t)$, where \mathbf{A}_0 is a constant vector, $\mathbf{k}(x, y, z)$ is a vector function of the coordinates constant with time. What is the expression of the electric field in that region?
- 10.8. A coaxial cable consists of two conducting coaxial cylindrical surfaces separated by a vacuum. The radii of the two conductors are R_1 and R_2 , with $R_1 < R_2$, as in Fig. 10.17. Suppose the resistivity of the conductors to be negligible.

Fig. 10.17 The coaxial cable
of problem 10.8



The circuit is closed with a resistor R at one end and a battery delivering the emf V at the other. The internal resistance of the battery is negligible. Find the Poynting vector between the two surfaces.

- 10.9. A capacitor of capacitance C discharges through the resistance R . Its charge varies with time as $Q = Q_0 \exp(-t/RC)$. How does the displacement current between the plates of the capacitor vary with time?
- 10.10. At point (x, y, z) , the current density is zero and the Poynting vector is $\mathbf{S} = y\mathbf{i} - x\mathbf{j} + a\mathbf{k}$ where a is constant. Is the energy density at that point constant or variable with time?

Index

A

Absolute magnetic permeability, 308
Acting field, 129, 133
Action and reaction law, 365
Action and reaction law violation, 365
Aharonov-Bohm solenoid effect, 391
Aharonov, Yakir, 391
Alternating current, 269
Ampere, 220
Ampère, 219
Ampère, André-Marie, 10, 181, 197, 201
Ampère circuital law, 201, 203
Ampère-Laplace law, 217
Ampère law, 319, 343
Ampere per meter, 302
Ampere-turns, 320, 326
Amplitude, 270
Angle sides, 33
Angle vertex, 34
Angular frequency, 270
Angular momentum, 314
Angular momentum conservation, 363, 368
Angular momentum density, 364
Arago, François, 259
Arnold, Harold, 31
Atlantic cable, 160
Atom, 3
Atomic number, 5
Atomic polarizability, 117
Axial vector, 222, 230
Ayrton, William, 356

B

Back electromotive force, 267
Ballistic galvanometer, 254
Bardeen, John, 173
Barkhausen effect, 331
Barkhausen, Heinrich, 329

Barkhausen jumps, 329
Barlow, Peter, 249
Barlow's wheel, 249
Base units, 149
Battery, 22
Bednorz, Johannes, 173
Bennet, Abraham, 3
Betatron, 250, 251
Bifilar line, 264
Biot, Jean-Baptiste, 197
Biot-Savart law, 198
Bohm, David, 391
Bohr model, 314
Bohr, Niels, 314
Boltzmann constant, 131, 318
Boltzmann factor, 318
Boltzmann, Ludwig, 131
Boundary energy, 328
British Association for the Advancement of Science, 161
British Association Ohm, 162, 350
Broken magnet experiment, 180, 313

C

Calutron, 191
Canonical momentum, 390
Capacitance of a capacitor, 81
Capacitive impedance, 274
Capacitor, 80, 167, 273, 278
Capacitor charging, 168
Capacitor discharging, 167
Capacitors connected in parallel, 86
Capacitors connected in series, 85
Capacity, 78
Capacity of a condenser, 81
Cavendish, Henry, 91
Cavity in dielectric, 134
Center of charge, 58

- Charge carrier, 63, 149
 Charge conservation, 154
 Charged cylindrical surface, 44
 Charge density, 14
 Charged plane, 46
 Charged sphere, 43
 Charged spherical shell, 41
 Charge invariance, 32, 232
 Circulation, 22
 Circulation integral, 41
 Clausius-Mossotti formula, 136
 Clausius, Rudolf, 136
 Clerk Maxwell, James, 228, 341
 Closed cavity, 72
 Coaxial cable, 265
 Coefficient of self induction, 262
 Coercive field, 321
 Coercivity, 321
 Color force, 3
 Complete induction, 80
 Complex amplitude, 270
 Complex notation, 271
 Condenser, 80
 Conduction current, 306
 Conduction current density, 372
 Conductor, 63
 Constant current, 269
 Continuity equation, 155, 341, 359
 Continuity equation for energy, 358
 Continuous voltage generators, 165
 Cooper, Leon, 173
 Cooper pairs, 173
 Correlation distance, 173
 Cosmic rays, 68
 Coulomb, 5, 10
 Coulomb, Charles, 180
 Coulomb, Charles Augustin, 7
 Coulomb experiment, 124
 Coulomb gauge, 210, 381, 388
 Coulomb's force, 10
 Coulomb's law, 9
 Counter-electromotive force, 267
 Covariance, 229, 230
 Curie, Pierre, 318, 331
 Curie temperature, 331
 Curie's law, 318
 Curl, 22
 Current density, 150
 Current density four-vector, 233
 Current intensity, 148, 149
 Cyclotron, 186
 Cylindrical capacitor, 84
 Cylindrical solenoid, 205
- D**
 Dalembertian, 385
 d'Alembert, Jean Baptiste, 385
 De magnete, 2
 Demagnetization, 320, 322
 Dempster, Arthur, 189
 Diamagnetic materials, 298, 310
 Diamagnetism, 316
 Dielectric constant, 85, 115, 132, 133, 136
 Dielectrics, 114
 Dielectric strength, 85
 Dipole, 40, 51
 Dipole approximation, 57
 Dipole magnets, 324
 Dipole moment, 287
 Direct current, 149, 269
 Discontinuity, 18
 Displacement current, 342, 343
 Domain theory, 328
 Double plane, 47
 Drift velocity, 153
- E**
 Earth magnetic field, 194
 Easy magnetization axis, 327, 328, 330
 Eddy currents, 258
 Effective value, 279
 Electric charge, 3
 Electric charge conservation, 7, 341
 Electric charge local conservation, 372
 Electric conductivity, 158
 Electric current, 148
 Electric dipole moment, 52, 57
 Electric displacement, 123, 342, 372
 Electric field, 13, 15, 39
 Electric field discontinuities, 48
 Electric pendulum, 125
 Electric permittivity, 10
 Electric resistance, 156
 Electric susceptibility, 117, 130, 376
 Electric work, 288, 289
 Electrodes, 23
 Electrolyte, 64
 Electromagnet, 323
 Electromagnetic force, 3
 Electromagnetic generator, 256
 Electromagnetic induction, 242
 Electromagnetic interaction, 236
 Electromagnetic potentials, 379
 Electromagnetic unit, 351
 Electromagnetic waves, 347, 349, 382
 Electrometer, 3
 Electromotive force, 165, 244

- Electron, 3
Electron mass, 6
Electronvolt, 5
Electroscope, 3, 67
Electrostatic capacitance, 78
Electrostatic field, 13
Electrostatic force, 10
Electrostatic induction, 66, 79
Electrostatic induction coefficient, 88
Electrostatic polarization, 117
Electrostatic potential, 20, 378
Electrostatic shield, 73, 89
Electrostatic spoon, 3
Electrostatic unit, 351
Elementary charge, 5, 27, 31
Emf generator, 272
Energy, 99, 140
Energy conservation, 363
Energy density, 104, 334, 359
Energy flux, 360, 369
Energy in a capacitor, 102
Energy of a capacitor, 102
Energy of a parallel-plate capacitor, 106
Energy of a point charge, 108
Energy of a spherical capacitor, 107
Energy of continuous charge distribution, 100
Energy of the current, 284
Energy of the electrostatic field, 104
Energy stored in an inductor, 284
Equality of electron and proton charge, 12
Equipotential surface, 40
Equivalent capacitance, 85
Equivalent circuit, 301
Equivalent dipole, 58
Event, 383
Explorer charge, 13
Exploring charge, 13
- F**
Farad, 11
Faraday cage, 90
Faraday law, 248, 391
Faraday, Michael, 78, 90, 114, 242
Felici, Riccardo, 253
Felici's law, 253, 320
Ferromagnetic materials, 298, 300, 311, 323, 327
Feynman, Richard, 367
Field angular momentum, 370
Field energy, 140
Field energy density, 140
Field momentum density, 364
Field of forces, 13
1st Laplace's law, 217
- Fizeau, Hippolyte, 350
Flat cavity, 134
Flip-coil method, 253, 320
Flux of a vector, 36
Flux of B, 244
Flux of electric field, 36
Flux rule, 244
Focusing power, 190
Foucault currents, 258
Foucault, Léon, 350
Four-current density, 384
Four-dimensional gradient, 385
Four-dimensional space, 383
Fourier integral, 377
Fourier, Jean Baptiste, 377
Fourier series, 377
Four-momentum, 384
Four-potential, 386
Four-vector, 383, 384
Franklin, Benjamin, 72
Free charge, 122, 123, 372
Free charges density, 372
Free electrons, 63
Frequency, 270
Fringe field, 139
Fringing effects, 82
- G**
Galilei, Galileo, 231, 382
Galilei transformations, 232, 382
Galvanometer, 151
Gases, 64
Gas thermometer, 172
Gauge fixing, 210, 380
Gauge invariance, 90, 380, 391
Gauge transformations, 380
Gauss, Carl Friederich, 38
Gauss law, 38, 41, 42, 44, 46, 48
Generator, 149
Gilbert, William, 2, 179
Good conductors, 156
Good insulators, 156
Gradient, 41
Gradient theorem, 21
Gravitational field, 13
Ground, 79
Guard ring, 82
- H**
Hall, Edwin, 184
Hall effect, 184
Hall probe, 185
Harmonic function, 51, 75
Helium, 6

Henry, 261, 262
 Henry, Joseph, 242, 261
 Hollow conductor, 72
 Hopkinson, John, 326
 Hopkinson's laws, 326
 Huges, Vernon W., 33
 Hydrogen, 6
 Hysteresis cycle, 321, 323, 325

I

Image charge, 92
 Impedance, 274
 Impedance of the free space, 346
 Induced emf, 244
 Induced moment, 59
 Inductance, 262
 Inductive impedance, 274
 Inductor, 272, 277, 323
 Initial phase, 270
 Insulators, 114
 Interaction energy, 99, 286, 290, 391–393
 Interaction momentum, 390, 393, 394
 Internal resistance, 166
 Intrinsic electric dipole moment, 58
 Inversion, 229
 Iron core, 323, 326
 Isolated conductor, 77
 Isotope, 5

J

Joule effect, 164, 361
 Joule, James Prescott, 164
 Joule's law of heating, 164

K

Kerst, Donald, 250
 Kirchhoff rules, 278
 Kirkhoff, Gustav, 170
 Kirkhoff rules, 170

L

Landé, Alfred, 316
 Landé factor, 316
 Laplace equation, 51, 74, 87, 91, 385
 Laplace, Pierre-Simon, 51, 183
 Laplacian, 50, 385
 Lawrence, Ernest, 186
 Lenz law, 247, 300
 Lenz, Heinrich, 247
 Light, 382
 Lighting rod, 72
 Lighting strike, 72
 Light velocity, 357
 Linear charge density, 15

Linear charge distribution, 16, 24
 Linear conditions, 117
 Linear dielectric, 117
 Linear magnetic materials, 334, 335
 Linear media, 375
 Linear momentum, 364, 384
 Linear momentum conservation, 363, 367
 Linear momentum density, 364
 Line of force, 39
 Linked magnetic flux, 244
 Local charge conservation, 155
 Local energy conservation, 357
 Long solenoid, 263
 Lorentz force, 182, 245
 Lorentz, Hendrik, 182
 Lorentz transformations, 232, 236, 382, 383
 Lorenz gauge, 381
 Lorenz, Ludwig, 381

M

Macroscopic charge, 227
 Macroscopic field, 65, 105, 300
 Macroscopic field energy, 332
 Magnetic charge, 313
 Magnetic charge density, 313, 324
 Magnetic circuit, 319, 324
 Magnetic dipole, 220, 286, 315
 Magnetic energy density, 293
 Magnetic field, 182
 Magnetic field discontinuities, 207
 Magnetic field energy, 290
 Magnetic flux, 326
 Magnetic focalization, 192
 Magnetic force, 182
 Magnetic history, 320, 325
 Magnetic mirror, 193
 Magnetic moment, 222, 314
 Magnetic monopole, 180
 Magnetic Ohm's laws, 326
 Magnetic pendulum, 298
 Magnetic permeability, 310, 323, 326
 Magnetic poles, 180, 312
 Magnetic reluctance, 326
 Magnetic spectrum, 197
 Magnetic storm, 195
 Magnetic susceptibility, 308, 309, 318, 376
 Magnetic walls, 328
 Magnetization, 301, 303, 305, 307, 311
 Magnetization current, 306
 Magnetization current density, 373
 Magnetization curve, 320, 325
 Magnetization density, 301, 373
 Magneton motive force, 319, 326
 Mass spectrograph, 189

- Mass spectrum, 189
Maxwell, 228, 236, 371
Maxwell equations, 341, 346, 371
Maxwell, James, 91, 346
Maxwell, James Clerk, 162, 198
Mechanical energy, 225, 288
Mechanical work, 287
Mendeleev, Dimitri, 4
Mendeleev table, 4, 5
Metals, 63
Method of images, 91
Michelson, Albert, 356
Micro-crystal, 327
Microscopic field, 105
Millikan, Robert, 27, 69
Molecular currents, 181
Molecular dipole, 131
Molecule, 3
Momentum density, 364
Momentum per unit charge, 209
Monopole, 51
Monopole approximation, 56
Mossotti, Ottaviano, 136
Müller, 173
Müller, Karl, 173
Mutual inductance, 261
Mutual induction, 260
Mutual induction coefficient, 261
- N**
Narrow cavity, 134
Negative charge, 3
Negative pole, 23
Neumann, Franz, 242
Neutron, 3
Nodes rule, 170
Nonpolar dielectric, 128, 129
Normal dielectric, 117, 123
n-type semiconductors, 63
Nucleon, 5
Nucleus, 3
Null method, 350
- O**
Ohm, 156
Ohm, Georg, 156
Ohm's law, 156, 158, 163
Ohm per square, 164
Ohmic conductor, 156
Oil-drop experiment, 27
Onnes, Kamerlingh, 171
Operational definition, 15
- Orbital angular momentum, 314
Orientation, 131
Ørsted, 197, 242
Ørsted. Hans Christian, 180
- P**
Pacini, Domenico, 68
Parallel-plate capacitor, 81
Paramagnetic materials, 298, 309
Paramagnetism, 318
Passive components, 272
Pelerin, Pierre, 180
Period, 270
Permanent magnet, 303, 313
Permanent magnetization, 311
Permeability of the free space, 198
Permittivity, 85, 115
Perpendicular component discontinuity, 48
Perry, John, 356
Phase, 270
Photon mass, 91
Pile, 149, 156
Planar charge distribution, 17, 25
Poincaré, Henri, 383
Poisson equation, 50
Poisson, Siméon Denis, 51
Polar dielectric, 128, 131
Polarizability volume, 130
Polarization, 117–119, 124, 126, 372
Polarization charge, 116, 123, 372
Polarization charge density, 121
Polarization current density, 373
Polarization density, 117
Poles, 23
Polycrystalline, 327
Polycrystalline structure, 328
Positive charge, 3
Positive pole, 23
Potential difference, 22
Potential linear momentum, 390, 391
Power, 164
Poynting, John Henry, 360
Poynting theorem, 360
Poynting vector, 360–362, 369
Priestley, Joseph, 90
Primary standard, 162
Proper energy, 108
Proper scalar, 230
Proper vector, 230
Proton, 3, 6
Pseudoscalar, 230
Psuedovector, 230

p-type semiconductor, 63
Pulsation, 270

Q

Quantum electrodynamics, 380
Quark, 3
Quasi-stationary regime, 155

R

Radian, 34
Randomization time, 158
Ratio of units experiment, 351
Reactance, 275
Receptor, 13
Relative magnetic permeability, 308
Relative permittivity, 85
Relativity principle, 183, 231, 236, 382
Residual resistance, 159
Resistivity, 156
Resistor, 272, 278
Resistors in parallel, 169
Resistors in series, 169
Resolving power, 190
Resonance, 277
Right hand rule, 181, 198, 231
Rms velocity, 153
Rotation, 229

S

Saddle point, 41
Savart, 197
Savart, Félix, 197
Scalar, 230
Scalar potential, 379
Schrieffer, 173
Schrieffer, John, 173
2nd Laplace's law, 183
Self-inductance, 262
Semiconductors, 63, 156
Shunt, 354
SI, 5, 149
Siemens, 158
Sink, 39
Système International, 356
Solar wind, 194
Solid angle, 34
Solid angle element, 36
Solid charged cylinder, 45
Source, 13, 39
Space-time, 383
Speed of light, 228, 232
Spherical capacitor, 83
Spherical cavity, 135
Spherical charged surfaces, 43

Spherical conductor, 78
Spin, 314, 315
Spinning coil, 162
Stable equilibrium, 76
Static field, 39
Steady regime, 155, 169
Steradian, 34
Stokes, George G., 31
Stokes law, 31
Strong nuclear force, 3
Superconductivity, 171
Superposition principle, 11, 14, 19, 47, 182
Surface charge density, 15
Surface charges, 225
Surface conductors, 163
Surface current density, 205
Surface currents, 163
Surface resistivity, 164
Synchrotron radiation, 252

T

Tangential component, 49
Thales from Miletus, 179
Thales of Miletus, 2
Thomson, Elihu, 258
Thomson, Joseph John, 27, 390
Thomson jumping ring, 258
Thomson, William, 82, 160, 350
Three-vector, 384
Time constant, 167, 268
Toroidal solenoid, 204, 263
Torsion balance, 7
Total impedance, 275
Transition temperature, 171

U

Unit cell, 327

V

Vacuum impedance, 346
Vacuum permeability, 198, 356
Vacuum permittivity, 10, 357
Van Allen belts, 193
Van Allen, James, 193
Van de Graaf generator, 149
Van de Graaf, Robert, 149
Vector, 230
Vector field, 39
Vector potential, 209, 252, 378, 388
Velocity of light, 350, 356
Velocity selector, 188
Virtual works principle, 54, 103
Volt, 21
Volta, Alessandro, 21, 23

- Voltage, 22
voltage generator, 166
Volta pile, 23
Volume charges, 225
Volume element, 36
- W**
Water, 131
Water molecule, 58
Wave equation, 349, 382
- Weak nuclear force, 7
Weber, 197
Weber, Wilhelm, 197
Weiss domains, 328
Weiss, Pierre-Ernest, 328
Work function, 106
- Z**
Zorn, Jens C., 33